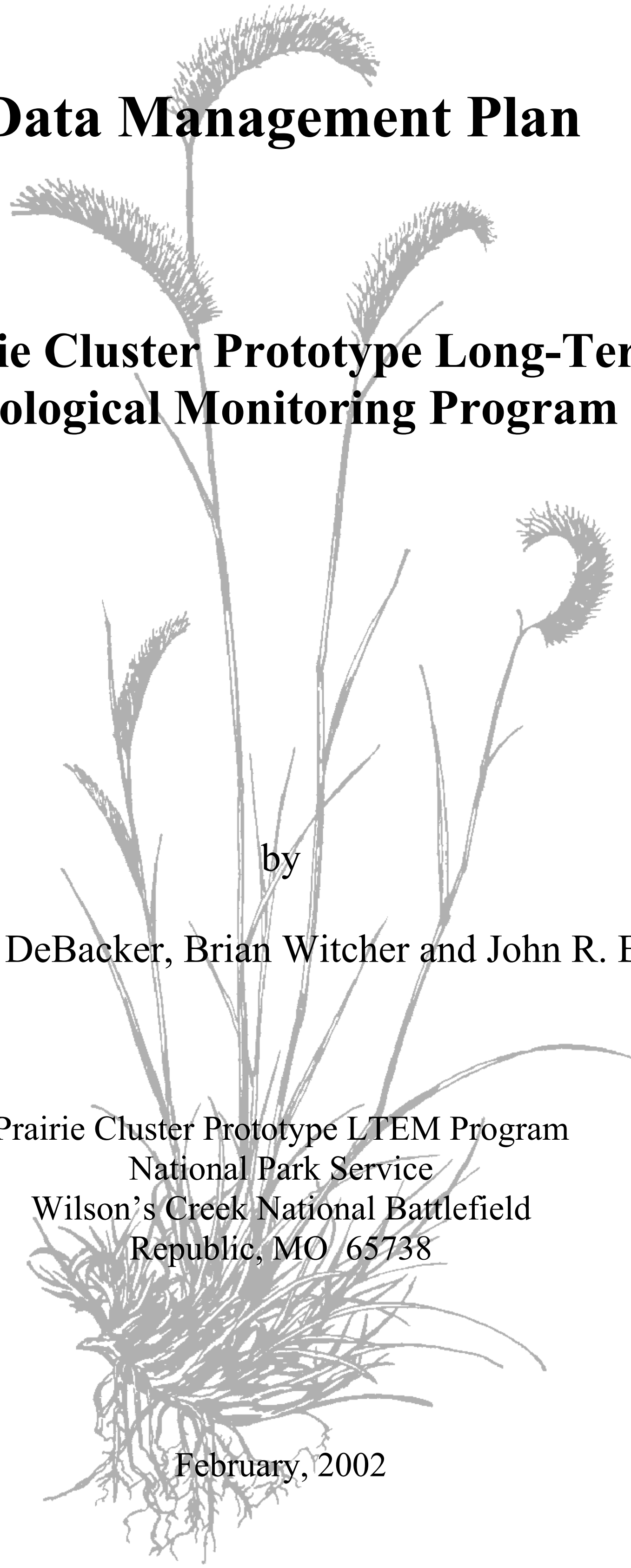


Data Management Plan

Prairie Cluster Prototype Long-Term Ecological Monitoring Program

by

Michael D. DeBacker, Brian Witcher and John R. Boetsch

Prairie Cluster Prototype LTEM Program
National Park Service
Wilson's Creek National Battlefield
Republic, MO 65738

February, 2002

Acknowledgements

We would like to thank the following people for their numerous contributions toward building our data management program and this document: Peter Budde, Donald Catanzaro, Linda Dye, Steve Fancy, Keith Grabner, Joe Gregson, Pat Lineback, Karen Oakley, Steve Tessler and Lisa Thomas.

TABLE OF CONTENTS

Description	Page
Introduction	1
I. Data Management Resources	2
A. Personnel	2
B. Computer Resources	4
C. Data Resources	6
II. Data Management	13
A. Database Infrastructure	13
B. Data Integration	15
C. Data Acquisition	16
D. Quality Assurance and Quality Control	17
E. Data Security	20
F. Data Summary, Analysis and Reporting	21
G. Data Posting and Dissemination	22
H. Data Management Roles and Responsibilities	22
III. Work Plan	24
Literature Cited	26
Appendices	
A. Description of Common Database Elements	A-1
1. Shared Lookup Tables	A-1
2. Core Metadata Tables	A-3
B. Plant Community Database Description	B-1
1. Database Structure	B-1
2. Data Entry	B-4
3. Data Summary and Reporting	B-5
C. Missouri Bladderpod Database Description	C-1
1. Database Structure	C-2
2. User Interface	C-10
3. Data Entry	C-11
4. Data Summary and Reporting	C-11
D. Grassland Birds Database Description	In development
E. State T&E Plants Database Description	In development
F. Western Prairie Fringed Orchid Database Description	In development
G. Black-tailed Prairie Dog Database Description	In development
H. Aquatic Macroinvertebrates Database Description	In development
I. Topeka Shiner Database Description	In development

LIST OF FIGURES

Figure Number	Description	Page
1	Hypothetical cost of assuring good data quality at various phases.	2
2	Computer resources and local area network.	5
3	Data types collected and managed – an example from grassland bird monitoring.	6
4	Data management process and resulting products.	13
5	Relationships among data tables – an example from the aquatic macroinvertebrate database.	14
6	Diagram of the data integration process.	16
7	QA/QC features of the plant community data entry form.	18
8	File structure for monitoring projects.	20

LIST OF TABLES

Table Number	Description	Page
1	Prairie Cluster staff resources directed towards data management.	3
2	Software packages currently in use.	5
3	Prairie Cluster monitoring protocols.	7
4	Summary of data collected for each monitoring project.	8
5	Base spatial data managed by the Prairie Cluster.	12
6	Annual schedule for data processing, reporting and archival.	21
7	Long-term schedule for data analysis and reporting.	22
8	Responsible parties for various aspects of data management.	23

INTRODUCTION

Initiated in 1994, the Prairie Cluster Prototype Long-Term Ecological Monitoring Program (Prairie Cluster Prototype Monitoring Program) monitors natural resources and resource issues within seven prairie parks that are widely scattered throughout the central grasslands. Program staff are based in one park (Wilson's Creek National Battlefield, Missouri), and collect monitoring data there and in six other parks (Agate Fossil Beds, Scotts Bluff, and Homestead National Monuments in Nebraska, Tallgrass Prairie National Preserve in Kansas, Pipestone National Monument in Minnesota and Effigy Mounds National Monument in Iowa). The principal functions of the program are the collection and analysis of data to detect changes in the health of park resources, and to develop techniques and strategies to improve long-term monitoring in these parks and throughout the National Park Service.

Our monitoring staff annually makes thousands of observations about plant and animal populations, communities and their environments. Taken together, these observations form a statistical representation of our sampling universe. In essence, the purpose of data management is to ensure that an accurate and complete record of those observations is maintained in perpetuity. Specific data management objectives relate to three themes: data management infrastructure, data integration and exploration, and long-term data integrity and security.

Data infrastructure

- Design a data management system comprised of MS Access databases that contain data tables, data entry forms, and summary reports.
- Implement a standardized file organization system.

Data integration and exploration

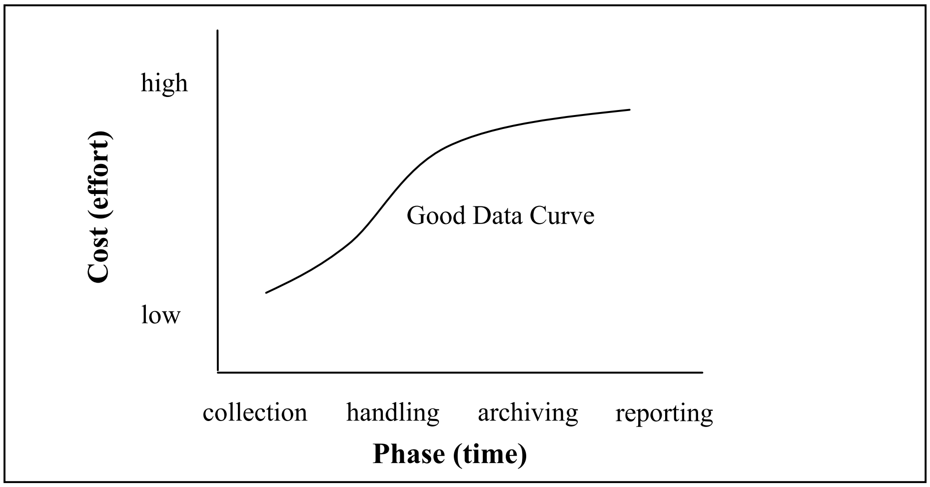
- Design common fields and tables to ensure compatibility among data sets.
- Create a user interface to easily explore data relationships, alter the period or spatial scale of interest, and answer questions about the data.

Long-term data integrity and security

- Utilize redundant data storage devices.
- Archive data and maintain an edit log.
- Create and maintain spatial and tabular metadata.

The focus of data management efforts within the Prairie Cluster Prototype Program is strategic within the context of service-wide data management activities. The Service-wide Inventory and Monitoring Program has invested considerable resources in developing tools to archive and disseminate data (e.g. NPSpecies, Dataset Catalog, Theme Manager). Further, general guidance regarding data management procedures for handling and validating data is contained in the Draft Data Management Protocol (Tessler & Gregson 1997). The Prairie Cluster data management system builds on these resources while emphasizing and expanding the role of data management during data collection and handling. Attention to data management early in the process targets resources where they are most efficient in affecting data quality (Figure 1).

Figure 1. The hypothetical cost of assuring good data quality at various phases.



The goal of the Data Management Plan is to describe the resources and process used to ensure high quality data. Specific objectives of the plan include the following:

- Describe the data management process.
- Describe roles and responsibilities of program staff for managing data.
- Describe the current hardware and software environment for managing data.
- Summarize data and associated metadata collected and/or managed by the Prairie Cluster and provide detailed descriptions of each data set in an appendix.
- Schedule routine summary reports and trend analysis reports.
- Define future direction of data management activities in a work plan.

While the Data Management Plan provides overarching data management guidance, project-specific data descriptions and procedures may also be found in the monitoring protocols. For each of the monitoring components (e.g. plant communities, grassland birds), the monitoring protocol details sample methods, summary routines and report format. Some of the protocols also include project-specific quality control procedures for conducting fieldwork. An overview of the Prairie Cluster program and our data management system is included in Thomas *et al*. (2001).

I. DATA MANAGEMENT RESOURCES

A. PERSONNEL

Assuring and maintaining data integrity is fundamental to the mission of a long-term monitoring program and requires a considerable investment of staff time. Table 1 gives an estimate of our staff resources committed to accomplishing data management.

Table 1. Prairie Cluster staff resources directed toward data management.

Position Title	# of Staff Positions	GS Grade	% of Time	Data Management Activities	Total FTE	Total Cost (k)
Program Coordinator	1	12	20%	analysis, data summary & reporting	0.2	16
Data Manager	1	9/11	75%	data archiving and dissemination, database development, report automation, assure overall QA/QC	0.75	49.5
Ecologists	3	9/11	30%	data validation, summary, analysis and reporting, oversee data entry and verification	0.9	59.5
Botanist	1	7/9	30%	data entry and verification	0.3	16.5
Bio-technician	1	5	30%	data entry	0.3	5.5
Program Total:					2.45	147

The Data Manager and Project Manager (lead ecologist on each monitoring project) share responsibility for data management within the Prairie Cluster organization. Typically, the Project Manager is responsible for data collection, data entry, verification and validation, as well as data summary, analysis and reporting. The Data Manager is responsible for data archiving, security, dissemination and database design. Furthermore, the Data Manager, in collaboration with the Project Manager, develops data entry forms and other database features to assure QA/QC and automate routine report generation. The Data Manager is ultimately responsible for developing adequate QA/QC procedures within the database management system and ensuring that appropriate data handling procedures are followed.

1. Data Manager

The Prairie Cluster Data Manager serves as GIS Specialist and Data Manager for the program. Many of the Data Manager's GIS duties include a substantial data management component. The principal data management responsibilities of the position are to:

1) Work with project managers to ensure that data sets are fully documented and validated. Responsible for metadata creation and maintenance.
2) Maintain archival copies of data sets and appropriate documentation.
3) Update data management aspects of project protocols in conjunction with project managers.
4) Write and maintain a program Data Management Plan.
5) Integrate tabular data with spatial data in a GIS system.
6) Maintain and update those elements of the local area network relevant to data management.
7) Provide basic training in the use of database software and service-wide data management tools.
8) Develop electronic data sharing and internet-based data dissemination.
9) Contribute to regional and national discussions regarding data standards and integration/analysis issues.

adapted from NPS Channel Islands Data Management Protocol (NPS 1998)

2. Project Manager

The primary data-related responsibilities of project managers are to:

1) Supervise and certify all field operations, including staff training, equipment calibration and data collection.
2) Maintain concise explanatory documentation of all deviations from procedures defined in the monitoring protocols.
3) Supervise or perform data entry, verification and validation.
4) Work with Data Manager to fully document and maintain master data.
5) Maintain hard-copy files of data and ensure copies are stored in a second location. Create timely trip reports referencing important details of each field data collection period.
6) Coordinate changes to database structure and field data forms with the Data Manager.
7) Work with Data Manager to create general MS Access tools (queries, reports) for annual reports and for users to access their data.
8) Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users.
9) Be the main point of contact concerning data content.

adapted from NPS Channel Islands Data Management Protocol (NPS 1998)

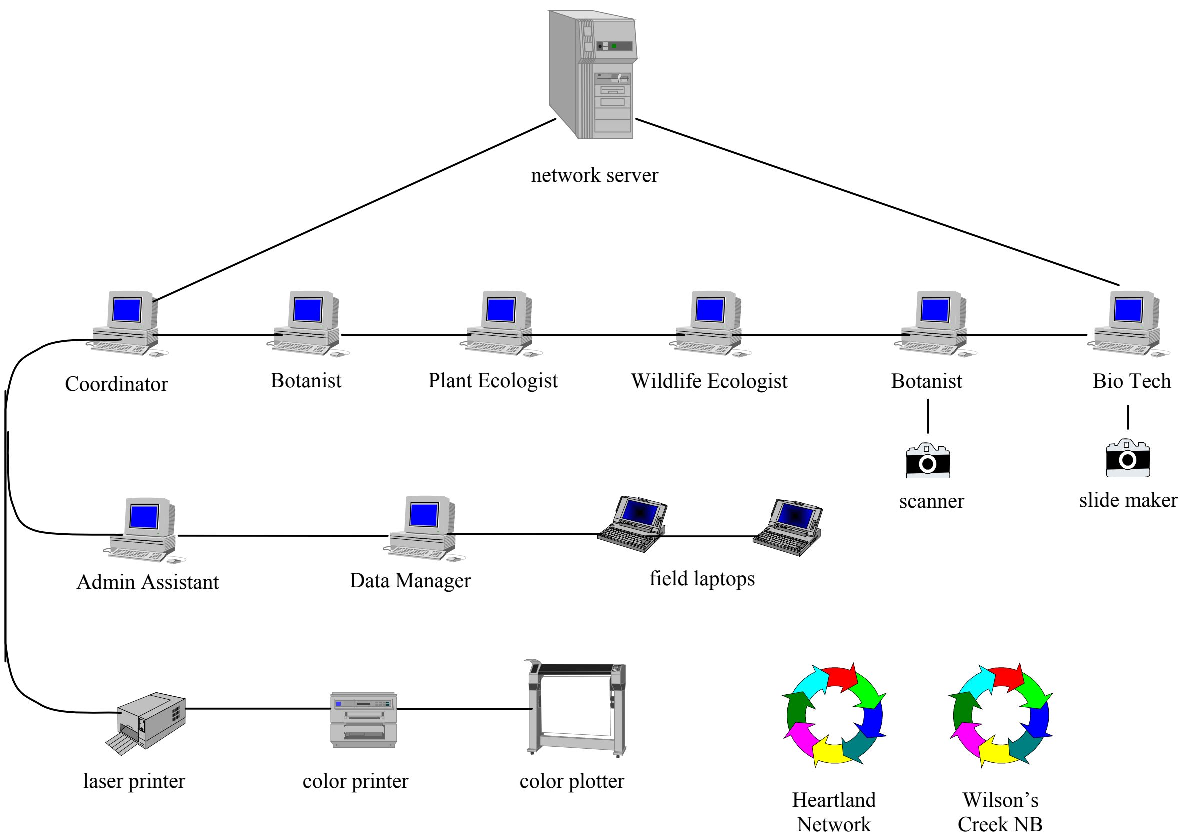
B. COMPUTER RESOURCES

1. Network, Hardware, Peripherals

Within the Prairie Cluster, all personal workstations are interconnected using a local area network (LAN). Through this LAN the Prairie Cluster is also connected to Wilson's Creek National Battlefield and the Heartland I&M Network. Use of a server allows for central data storage and management of data set access. The server's six hard drives have a Raid 5 configuration (Redundant Array of Inexpensive Disks). The Raid 5 configuration assures data sets are duplicated at all times on two different drives.

Current computer resources are 8 personal workstations, one GIS machine, four field laptops and one server. Peripherals attached to the network include a laser printer, small format color printer, large format color plotter, flatbed scanner, and a slide maker (Figure 2). Internet access is maintained through a frame relay connection to Midwest Regional offices in Omaha. A central router handles all internet traffic via this frame relay connection. Each workstation is connected to a battery backup unit to protect against power surges and outages. These units allow users to save their work and shut down in a normal manner in the event of a sustained power outage.

Figure 2. Computer resources and local area network.



2. Software Applications

Table 2 provides a list of the software packages currently in use within the Prairie Cluster Prototype Program. For the most part, we employ the service-wide standard software packages and versions.

Table 2. Software packages currently in use.

APPLICATION/FUNCTION	SOFTWARE PACKAGE/VERSION
Network Operating System	Microsoft NT Server 4.0
Tape Backup	ARCserve for NT
Environment Operating System	Microsoft Windows NT 4.0 and Windows 2000
Database	Microsoft Access 97
Word Processing	Microsoft Word 97
Spreadsheet	Microsoft Excel 97
Presentation	Microsoft PowerPoint 97
Graphics	Corel Draw 97, Axum
Digital Images	Camedia, Jasc Paint Shop, Adobe Photoshop
Bibliographic	Procite 4
GIS Desktop	ArcView 3.2, ArcInfo 8.1
Scheduling	Microsoft Outlook, Lotus Notes 5.0
Statistics	NCSS 2000, Pass 6.0, PC-Ord 4.0, SAS 8.2
Windows Utilities	WinZip 7.0
File Transfer	WS FTP95 LE
Virus Protection	Inoculate-IT

C. DATA RESOURCES

For each monitoring project, the monitoring data set is actually composed of many separate yet complimentary data sets (Figure 3). The following categories describe and assist organization of the various data managed by a long-term monitoring program. The monitoring staff directly collects primary data, metadata and habitat/environmental data.

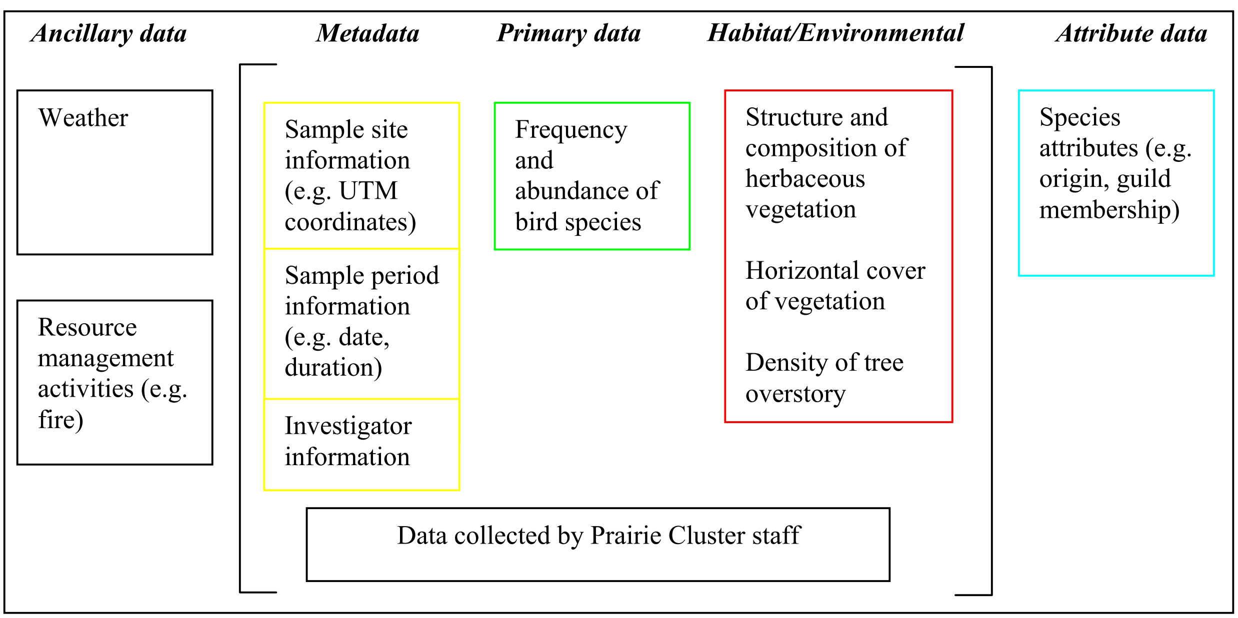
Primary data -- Direct measures of the population, community, or resource that is the focus of the monitoring protocol (e.g. plant species occurrence and abundance for plant community monitoring; prairie dog counts for prairie dog monitoring).

Metadata -- Standardized data describing the where, when and who of primary data collection.

Habitat/ environmental data -- Protocol-specific data tables describing the physical, chemical or biotic aspects of the habitat (e.g. water temperature, stream flow and substrate description for macroinvertebrate sampling; weather conditions and vegetation structure for bird monitoring).

Ancillary/ attribute data -- Management actions (e.g. prescribed fire, herbicide application) and climate (e.g. precipitation) are examples of ancillary data that are applicable to all project areas. Attribute data, such as a bird species nesting guild, are typically protocol-specific. Attribute and ancillary data are generally not collected by staff, but instead are acquired from third parties.

Figure 3. Data types collected and managed – an example from grassland bird monitoring.



1. Prototype Monitoring Program Project Data

Scientists with the USGS Biological Resource Division, university scientists, and Prairie Cluster staff have developed the project protocols. Each protocol has undergone a review process to ensure that the methods and analysis procedures provide scientifically sound data. Table 3 lists the protocols that constitute the Prairie Cluster monitoring program. Table 4 summarizes the type of data collected for each project.

Table 3. Prairie Cluster monitoring protocols.

Terrestrial Ecosystem

Landscape Monitoring

1. Adjacent land use

Community Monitoring

2. Grassland plant communities
3. Grassland birds
4. Grassland butterflies

Population Monitoring

5. State-listed T&E plants
6. Missouri bladderpod (*Lesquerella filiformis*)
7. Western prairie fringed orchid (*Platanthera praeclara*)
8. Black-tailed prairie dog (*Cynomis ludovicianus*)

Environmental Monitoring

9. Local climate

Aquatic Ecosystem

Community Monitoring

10. Macroinvertebrates as indicators of stream health

Population Monitoring

11. Topeka shiner (*Notropis topeka*)

Information is synthesized or derived from data. Consequently, the source and quality of underlying data is important when information is used to make long-term decisions. A lot of information may be based on good data, but much is not. . . . Collecting and managing data is a serious business. – Steve Tessler, 1995 (from Tessler & Gregson 1997)

Table 4. Summary of data collected for each monitoring project.

<table>
<tr><td>Title:
Adjacent Land Use</td><td>Protocol: In preparation</td><td>Project Manager:
Data Manager
(Brian Witcher)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Digital ortho photos from historic and current time periods.

<u>Habitat/Environmental data:</u>

<u>Status:</u>
In development: AGFO, EFMO, HOME, PIPE, SCBL, WICR</td></tr>
<tr><td>Title:
Grassland Plant Communities</td><td>Protocol: Willson, G.D., L.P. Thomas, M.D. DeBacker, W.M. Rizzo and C. Buck. 2001. Plant community monitoring protocol for six prairie parks. Biological Resources Division, U.S. Geological Survey, prepared for Great Plains Prairie Cluster Long-Term Ecological Monitoring Program, Republic, MO.</td><td>Project Manager:
Botanist
(Mike DeBacker)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Frequency and foliar cover of herbaceous and shrub species; number and size of woody species; density of seedlings and saplings.

<u>Habitat/Environmental data:</u>
Slope and aspect of sample sites, ground cover of bare soil, exposed rock, leaf litter, etc.

<u>Status:</u>
Implemented at AGFO, EFMO, HOME, PIPE, SCBL, TAPR, WICR</td></tr>
<tr><td>Title:
Grassland Birds</td><td>Peitz, D.G. and S.G. Fancy. In preparation. Bird monitoring protocol for Agate Fossil Beds National Monument, Nebraska and Tallgrass Prairie National Preserve, Kansas. Prairie Cluster Prototype LTEM Program, National Park Service, Republic, MO.</td><td>Project Manager:
Wildlife Ecologist
(David Peitz)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Frequency and abundance of bird species.

<u>Habitat/Environmental data:</u>
Water cover, composition and structure of plant community, horizontal cover of vegetation, density of tree overstory.
Temperature, wind speed, rain, cloud cover, and noise level.

<u>Status:</u>
In development, initiated at AGFO, TAPR</td></tr>
</table>

Table 4. Summary of data collected for each project (continued).

<table>
<tr><td>Title:
Grassland Butterflies</td><td>Protocol: Debinski, D., S. Mahady, W.M. Rizzo, and G.D. Willson. 2000. Butterfly monitoring protocol for four prairie parks. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 25 p.</td><td>Project Manager:
Wildlife Ecologist (David Peitz)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Frequency and abundance of butterfly species.

<u>Habitat/Environmental data:</u>
Temperature, wind speed, wind gust, cloud cover.

<u>Status:</u>
Pilot project conducted 1997 and 1998 at EFMO, HOME, PIPE, WICR</td></tr>
<tr><td>Title:
State-listed Threatened and Endangered Plants</td><td>Protocol: DeBacker, M.D., L.P. Thomas and J.R. Boetsch. In preparation. A practical framework for monitoring rare plant species. Prairie Cluster Prototype LTEM Program, National Park Service, Republic, MO.</td><td>Project Manager:
Plant Ecologist (John Boetsch)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Distribution, persistence and size of rare plant populations.

<u>Habitat/Environmental data:</u>
Habitat characteristics (slope, aspect, plant community).
Threat assessment (evidence of impacts to population or to the habitat).

<u>Status:</u>
In development, initiated at AGFO, SCBL, EFMO, PIPE, WICR</td></tr>
<tr><td>Title:
Missouri Bladderpod (Lesquerella filiformis)</td><td>Protocol: Kelrick, M.I. 2001. Missouri bladder-pod monitoring protocol for Wilson's Creek National Battlefield. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 28 p.</td><td>Project Manager:
Plant Ecologist (John Boetsch)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Abundance data used to estimate population size, population demographic data (e.g. individual survivorship and reproductive output).

<u>Habitat/Environmental data:</u>
Slope and aspect, litter and soil depths of demography plots. Estimated cover of associated species, bare soil, exposed rock, leaf litter, etc. See also Local Climate.

<u>Status:</u>
Implemented at WICR</td></tr>
</table>

Table 4. Summary of data collected for each project (continued).

Title: Western Prairie Fringed Orchid (*Platanthera praeclara)*	Protocol: Willson, G.D. 2001. Western prairie fringed orchid monitoring protocol for Pipestone National Monument. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 18 p.	*Project Manager*: Plant Ecologist (John Boetsch)
Primary data sets: Distribution, plant height and numbers of flowers of reproductive individuals. Density of non-flowering individuals. *Habitat/Environmental data:* See Local Climate. *Status:* Implemented at PIPE		
Title: Black-tailed Prairie Dog (*Cynomis ludovicianus*)	Protocol: Plumb, G. E., G. D. Willson, K. Kalin, K. Shinn, W.M. Rizzo. 2001. Black-tailed prairie dog monitoring protocol for seven prairie parks. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 27 p.	*Project Manager*: Wildlife Ecologist (David Peitz)
Primary data sets: Population counts and colony extent. *Habitat/Environmental data:* Temperature, wind speed, wind direction, cloud cover, precipitation. Density of colony vegetation. *Status:* Implemented at SCBL		
Title: Local Climate	Protocol: Akyuz, F.A. and P. Guinan. 2000. Weather monitoring protocol for two prairie parks. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 27 p.	*Project Manager*: Data Manager (Brian Witcher)
Primary data sets: Air and soil temperature, wind speed and direction, relative humidity, average solar radiation, fuel moisture and soil moisture. *Habitat/Environmental data:* *Status:* Implemented at PIPE and WICR		

Table 4. Summary of data collected for each project (continued).

<table>
<tr><td>Title:
Macroinvertebrates as Indicators of Stream Health</td><td>Protocol: Peterson, J.T., W.M. Rizzo, E.D. Schneider, and G.D. Willson. 1999. Macroinvertebrate biomonitoring protocol for four prairie streams. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 46 p.</td><td>Project Manager:
Wildlife Ecologist (David Peitz)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Frequency and abundance of aquatic macroinvertebrate species.

<u>Habitat/Environmental data:</u>
Stream conditions (e.g. temperature, depth, vertical displacement, substrate composition, gauge height).

<u>Status:</u>
Implemented at AGFO, HOME, PIPE, WICR</td></tr>
<tr><td>Title:
Topeka Shiner (Notropis topeka)</td><td>Protocol: Peitz, D.G. In preparation. Long-term monitoring protocol for Topeka shiner (Notropis topeka) in National Park Service Units within the Midwest Region, with emphasis on Tallgrass Prairie National Preserve, Kansas and Pipestone National Monument, Minnesota. Prairie Cluster Prototype LTEM Program, National Park Service, Republic, MO.</td><td>Project Manager:
Wildlife Ecologist (David Peitz)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Distribution and abundance of Topeka Shiner.
Species composition and relative abundance of fish community.

<u>Habitat/Environmental data:</u>
Air temperature, water temperature, weather conditions, turbidity, dissolved oxygen, off-channel pools, in stream flow, dominant substrate, substrate stability, stream bank stability, adjacent vegetation, pool size.

<u>Status:</u>
In development, initiated at PIPE and TAPR</td></tr>
</table>

2. Base Spatial Data

A variety of spatial data is important for long-term monitoring. Spatial data are necessary for selecting sample sites and useful as a correlate with monitoring results. All base spatial data are maintained on the server in a read-only format. All versions of spatial data considered to be working data sets are maintained in a separate directory. Table 5 summarizes the spatial data maintained by the Prairie Cluster.

Table 5. Base spatial data managed by the Prairie Cluster.

GIS Layer	AGFO	EFMO	HOME	PIPE	SCBL	TAPR	WICR
Boundary	X	X	X	X	X	X	X
Hydrology	X	X	X	X	X	X	X
Hypsography		X				X	X
Elevation	X	X	X	X	X	X	X
Vegetation	X_3	X_2	X_2	X_2	X_3	X_1	X_2
Soils	X	X	X	X	X	X	X
Trails	X	X		X	X	X	X
Roads	X	X	X	X	X	X	X
Structures							
DOQQs	X	X	X	X	X		X
Digitized USGS Maps		X	X	X		X	X
Prescribed Fire							X
T & E Species	X			X	X		X
Exotics				X			
Wetlands		X	X	X		X	

X_1 Based on 30m resolution satellite data
X_2 Based on management unit maps
X_3 USGS vegetation maps

3. Legacy Data

Monitoring of the biological world is not a new endeavor and many data sets are available for parks. Unfortunately, much of the existing data is poorly documented and not verifiable, and as such has little value. However, in some cases sufficient documentation exists to make the data potentially useful, and the Prairie Cluster is organizing and cataloging these data. Wherever possible these data sets are checked for consistency and accuracy. Legacy data can be defined as data collected prior to protocol implementation and employing different methods than those stipulated in the protocol. Protocol descriptions in Appendices B-I list available legacy data.

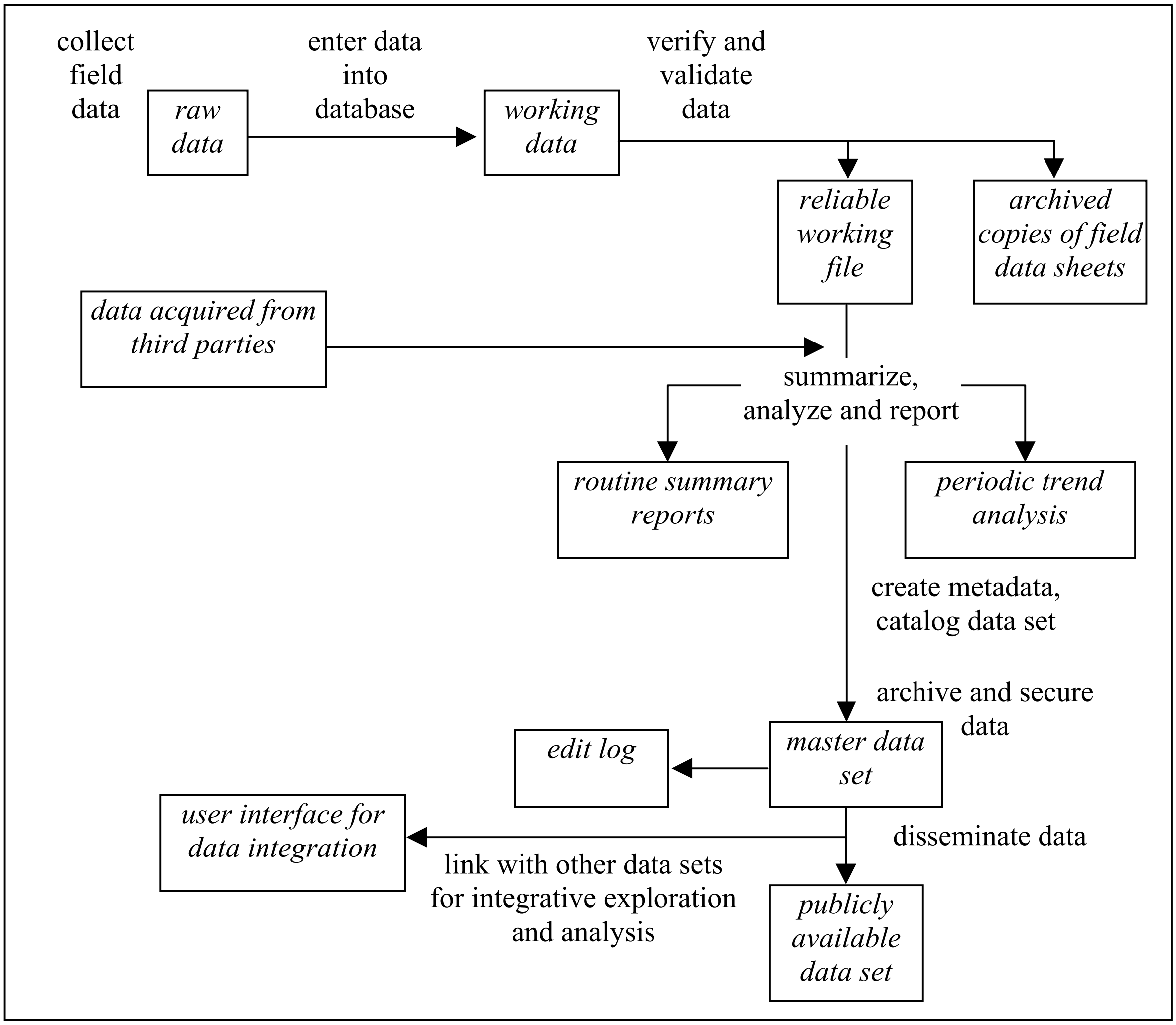
4. Data Organization Tools

The Prairie Cluster uses a number of tools provided by the Washington I&M Support Office (WASO I&M) to organize and disseminate data. For example, the Dataset Catalog is used to record data sets generated by the Prairie Cluster. Spatial data are packaged with the Theme Manager and distributed to Prairie Cluster parks. NPSpecies is used to record species observations and track source reports and voucher information. Finally, all reports and publications relevant to Prairie Cluster parks are entered into NPBIB. Many of these tools can be accessed through the I&M webpage: http://www1.nature.nps.gov/im/apps/index.htm.

II. DATA MANAGEMENT

The Prairie Cluster's strategy for data management incorporates database management tools, data handling procedures and supporting documentation to maintain high quality data. Figure 4 describes the process of data management from data collection to data archival and dissemination.

Figure 4. The data management process and resulting products.

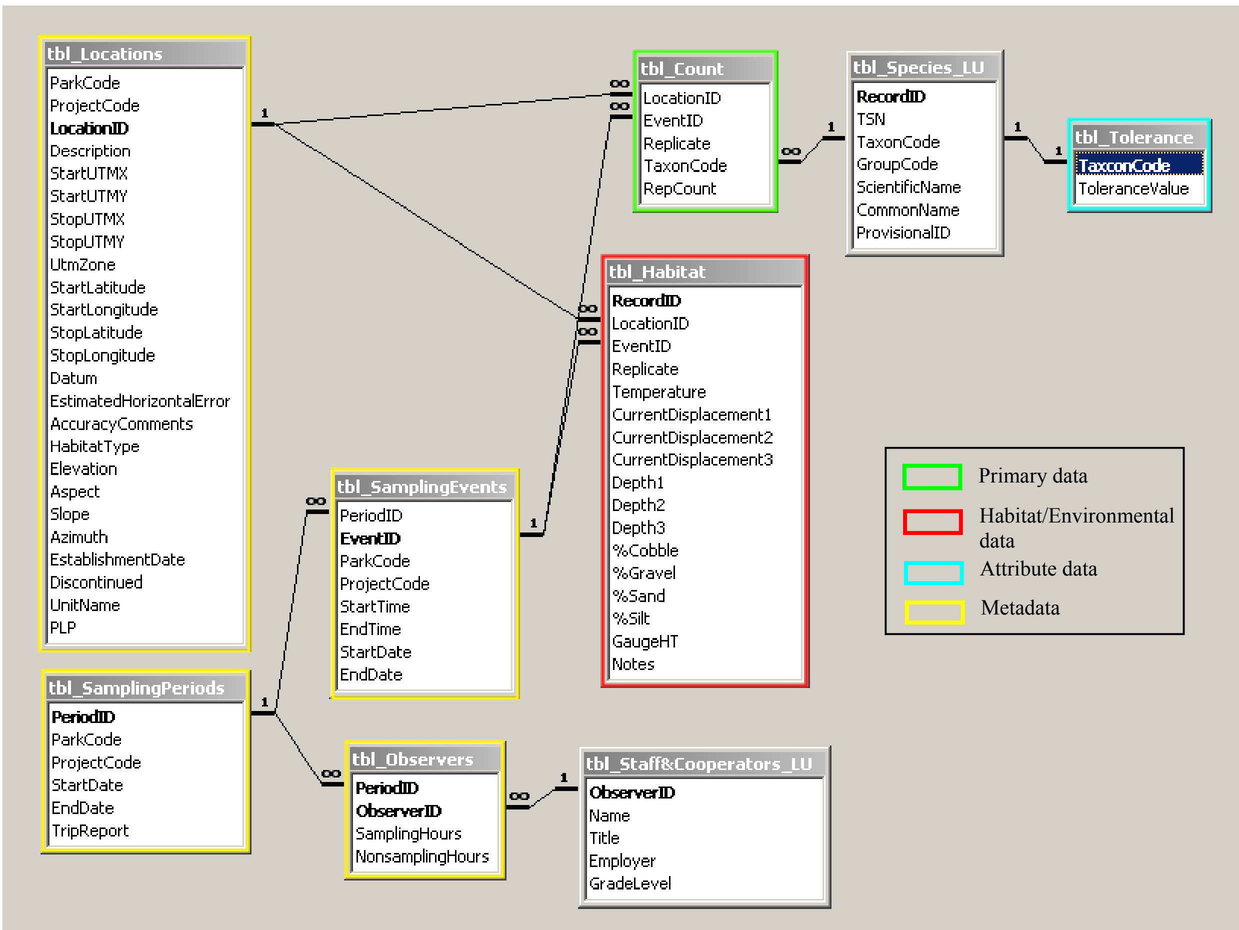


A. DATABASE INFRASTRUCTURE

The relational database system in use by the Prairie Cluster accommodates a diversity of data, eliminates redundancy, and maintains integrity among data tables through the use of software tools. Our database infrastructure is composed of multiple project databases developed and managed in MS Access. The basic design is modeled after work done at Channel Islands, with modifications to incorporate service-wide data standards. The overall database system is modular; that is, each project database functions as a stand-alone, but common elements in each database ensure that data can be combined for integrative reporting and analysis.

The database for each monitoring protocol contains a series of related data tables. Figure 5 shows the tables and relationships that constitute the database for aquatic macroinvertebrate monitoring. In this example, the occurrence of macroinvertebrate species (tbl_Count) is linked to essential metadata regarding the time and place of data collection. Macroinvertebrate occurrence data are also related to species attributes (e.g. tolerance values) and habitat/environmental data. Relationships among tables, indicated in Figure 5 by lines, are maintained by the software system and ensure integrity among data tables. For example, the database system requires that fields in the location table be properly populated before the location can be associated with observation data. Appendices B-I provide detailed descriptions of each project area database.

Figure 5. Relationships among data tables – an example from the aquatic macroinvertebrate database.



Through the sharing of common lookup tables and core data tables, all data within the prototype program are interrelated and can be readily exported to service-wide data sets. These common database elements are described briefly below, with more detail provided in Appendix A.

All protocol databases have in common a set of core metadata tables containing information that describes the time and place of data collection. Core metadata tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized across project databases with regard to field names, length and data type, and are compatible with the service-wide standard data fields proposed by WASO I&M. We have three core metadata tables common to each project database: locations, sampling periods, and observers tables. The locations table describes the site of data collection – its georeferenced position, the accuracy of georeferencing data, the date it was established, etc. The sampling periods table describes the period of time – typically on the order of days or weeks – during which the collective suite of sample sites that compose the statistical sample was visited to collect data. The observers table contains information about which observers were involved in a particular sampling period. A fourth table, sampling events, is optional but also standardized across projects. This table describes a much narrower time scale – typically on the order of a few hours – that is nested within a given sampling period. This temporal nesting was developed to accommodate projects where temporally variable conditions such as local temperature and weather can have a significant influence on the resources being monitored (e.g., birds often don't call while it's raining). One sample period can have several sampling events nested within it.

In contrast, common lookup tables are stored in a central location; each of the project databases refers to these central lookup tables rather than storing redundant information locally. Examples of lookup tables include: parks metadata, project administration and management information, species nomenclature and attribute information, the cover classes used to estimate percent cover in a defined area, and information on the various staff members and field observers. The use of common lookup tables helps to ensure a basic level of cross-project compatibility by encouraging the use of standard field formats and names wherever possible.

B. DATA INTEGRATION

The capability to query information across many project areas is a powerful analytical tool. Interdisciplinary and regional data exploration adds value to the data as it has the potential to reveal larger patterns in resource trends and can build a stronger body of evidence for overall resource degradation. Because our database system is designed around stand-alone databases for each project area, we have emphasized a multi-source, or federated, approach to data integration. This approach takes advantage of the fact that the data sets are modular, and only integrates data from these distributed sources on demand. The use of core data tables and common lookup tables, as described in the previous section, facilitates compatibility across project areas.

Our mechanism for integration is a front-end graphic user interface (GUI) that presents the user with views and descriptions of the raw and summarized data available for each project. Based on user input, batch commands and queries automatically process and retrieve information as needed from multiple back-end data sets, delivering and integrating it as if it were all locally stored (Figure 6). Data can then be explored and summarized in an interdisciplinary manner, or be combined with other integrated data sets at the regional or national level. At the end of the session, the user can choose to retain the versioned data if they were in the middle of an in-depth analysis. Otherwise the data are dumped and refreshed with every new session, from the relatively stable archived data sets, to avoid version control issues. Copies of the GUI can also be distributed so that multiple users can work independently on integrative analysis.

This virtual data centralization means that the front end consists primarily of data definitions, mapped paths to the distributed data sets, and batch commands to extract and transform data for integration. In this way, data are not duplicated and stored centrally except where they are being retained for active integrative exploration and analysis. Instead, the individual project data sets are maintained and archived separately. There are numerous advantages to this federated approach to data integration:

Advantages of a federated approach to data integration:

- Data are integrated only on demand, and are retrieved from the most recent archived copy of each project data set. This avoids version control problems related to comparing and updating centralized data from a local data set.

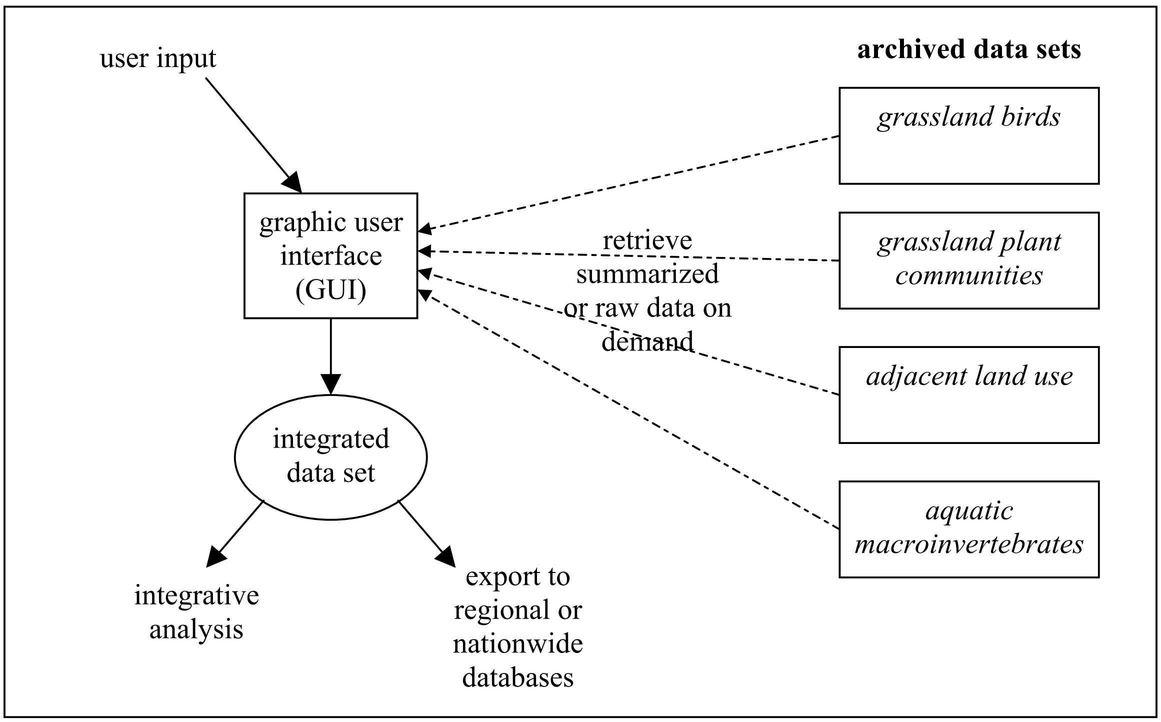
- Data sets are modular, allowing greater flexibility in accommodating the needs of each project area. A single data model does not therefore dictate the methods or kind of project data that are collected.

- Data summaries are optimized locally and exported for integrative analysis, which ensures consistency in information content because the integrative user isn't forced to interpret and manipulate raw data unless they choose to do so.

- Individual project databases and protocols can be developed at different rates without a significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.

- By working up from modular data sets, we avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very different – and often unforeseen – structural requirements. Furthermore, the payoff for this initial investment may not be realized down the road by greater efficiency for interdisciplinary use.

Figure 6. Diagram of the data integration process.



C. DATA ACQUISITION

For the most part, Prairie Cluster staff collect the monitoring data. Occasionally, park resource personnel collect a portion of the data. If data are to be collected by a third party (e.g. contractor, cooperator), then issues of data ownership should be addressed. Guidance for establishing data ownership is outlined in the Draft Data Management Protocol (Tessler & Gregson 1997). The Project Manager is responsible for data collection, and for implementing any QA/QC procedures specific to the sampling protocol. It is beyond the scope of this document to review the sampling methods of each protocol; rather, the following are some general program goals for data management during the data collection phase. First, each protocol should have specific QA/QC procedures for sampling. For example, the Plant Community Monitoring Protocol specifies alternating among sampling partners to avoid divergent estimates of foliar cover and describes a process for recording and tracking unknown specimens. Second, field forms should be concise, include units of measure, and prompt the user to collect all relevant data while promoting efficient data collection. Finally, each sampling period should be described in a trip report. The report emphasizes such things as weather conditions, logistical problems and any subsequent departure from the protocol, species identification problems, etc.

D. QUALITY ASSURANCE AND QUALITY CONTROL

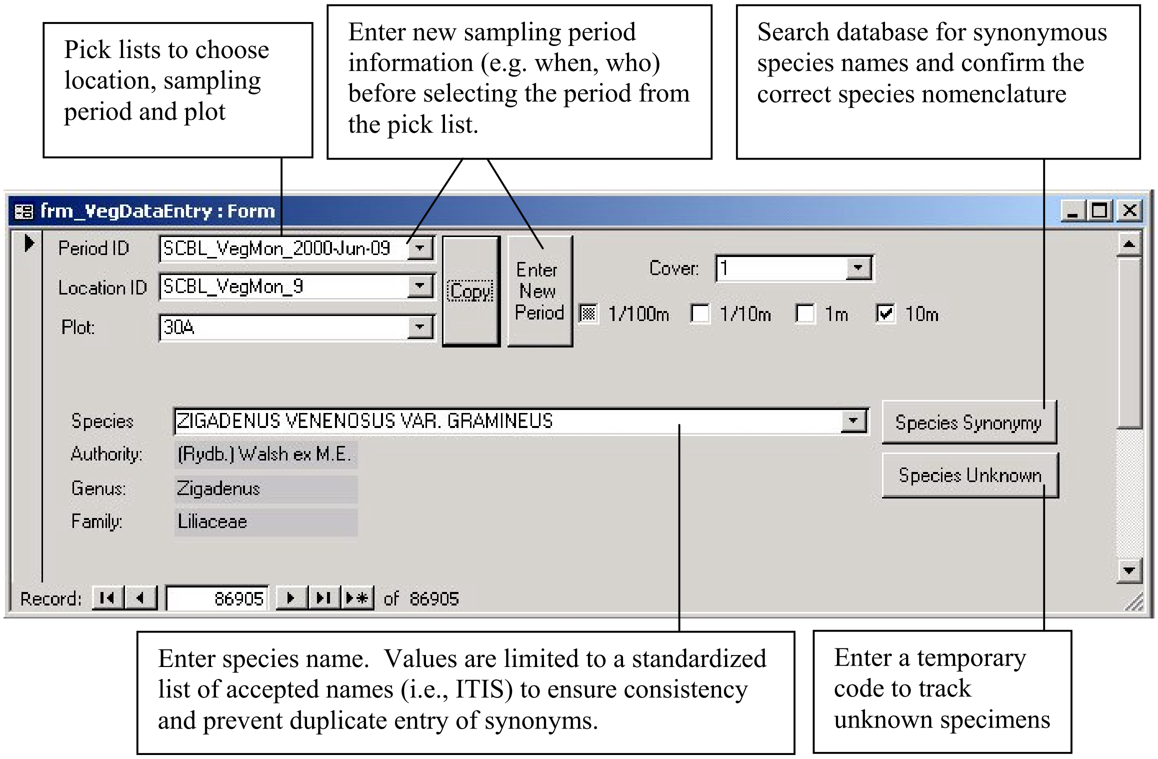
1. Data Handling Procedures

While an appropriate database infrastructure improves data documentation, integration, and integrity, good handling procedures protect initial data quality. During the data handling phase, data management activities focus on: 1) designing tools for data entry that reduce transcription errors; 2) independently verifying data transcription; and 3) developing data validation techniques. General procedures for quality assurance and control follow the Draft Data Management Protocol (Tessler & Gregson 1997) and the Channel Islands DMP (NPS 1998). Where necessary, additional features and procedures have been developed by the Prairie Cluster program and implemented for specific projects. A summary of these procedures follows, including excerpts from Tessler & Gregson (1997). Emphasis has been given to the additions and modifications made by the Prairie Cluster.

Data Entry

Data entry is the initial set of operations in which raw data from paper field forms or field notebooks are transcribed into a computerized form (i.e., within a database). Data entry is best performed by a person who is familiar with the data, and ideally takes place as soon as data collection is complete. Inevitably, the process of transcribing data from field forms to a digital format introduces error. However, data entry forms and QA/QC features have been developed for each protocol to minimize error. For example, key fields are set to prevent duplicate entry of data. Data entry forms reduce transcription errors through pick lists and value limits and provide controlled access to the database (i.e. forms are set for data entry only which prevents accidental deletion or alteration of existing data). Forms also control the sequence of data entry. For example, synonymous names (i.e. two or more different names referring to the same taxon) are common for plant species. Through the data entry form, a user searches for synonymous names before entering a new species name, thereby preventing the duplicate entry of synonyms. Figure 7 demonstrates some of the QA/QC features of the plant community data entry form.

Figure 7. QA/QC features of the plant community data entry form.



Data Verification

Data verification immediately follows data entry and involves checking the accuracy of computerized records against the original source—usually paper field records. While the goal of data entry is to achieve 100% correct entries, this is rarely accomplished. To minimize transcription errors, our policy is to verify 100% of records to their original source by permanent staff. Further, 10% of records are reviewed a second time by the Project Manager and the results of that comparison reported with the data. If errors are found in the Project Manager's review, then the entire data set is verified again. Once the computerized data are verified as accurately reflecting the original field data, the paper forms are archived and the electronic version is used for all subsequent data activities.

Data Validation

Although data may be correctly transcribed from the original field forms, they may not be accurate or logical. For example, a stream pH of 25.0 or a temperature of 95°C is illogical and almost certainly incorrect, whether or not it was properly transcribed from field forms. The process of reviewing computerized data for range and logic errors is the validation stage. Certain components of data validation are built into data entry forms (e.g. range limits). Additional data validation can be accomplished during verification, if the operator is sufficiently knowledgeable about the data. The Project Manager will validate the data after verification is complete. Validation procedures seek to identify generic errors (e.g. missing, mismatched or duplicate records) as well as errors specific to particular projects. For example, validation of plant community data includes database query and comparison of data among years. One query detects records with a location ID from a park and a period ID from a different park. Another query counts the number of plots per sample site (typically there are 10) to assure that all plots were entered. Finally, data are compared to previous years to identify gross differences. For example, *Dichanthelium oligosanthes* may be recorded this year, but *Dichanthelium spp* the previous.

During the entry, verification and validation phases, the Project Manager is responsible for the data. The Project Manager must assure consistency between field forms and the database by noting how and why any changes were made to the data on the original field forms. In general, changes made to the field forms should not be made via erasure, but rather through marginal notes or attached explanations. Once validation is complete, the data set is turned over to the Data Manager for archiving and storage.

Data Archival and Storage

Table 6 contains the schedule for data archival. Once the data are archived, any changes made to the data must be documented in an edit log. At this point forward, original field forms are not altered. Field forms can be reconciled to the database through the use of the edit log.

Secure data archiving is essential for protecting data files from corruption. Once a data set has passed the QA/QC procedures specified in the protocol, a formal entry is made in the I&M Data Set Catalog. Subsequently, an electronic version of the data set is maintained in a read-only format on the program server. Backup copies of the data are maintained at the Wilson's Creek visitor center, and an additional digital copy is forwarded to the NPS Inventory and Monitoring Program Archive.

Data Maintenance

Data sets are rarely static. They often change through additions, corrections, and improvements made following the archival of a data set. There are three main caveats to this process:

1) Only make changes that improve or update the data while maintaining data integrity.
2) Once archived, document any changes made to the data set.
3) Be prepared to recover from mistakes made during editing.

Any editing of archived data is accomplished jointly by the Project Manager and Data Manager. Every change must be documented in the edit log and accompanied by an explanation that includes pre- and post-edit data descriptions. The reader is referred to Tessler & Gregson (1997) for a complete description of prescribed data editing procedures and an example edit log.

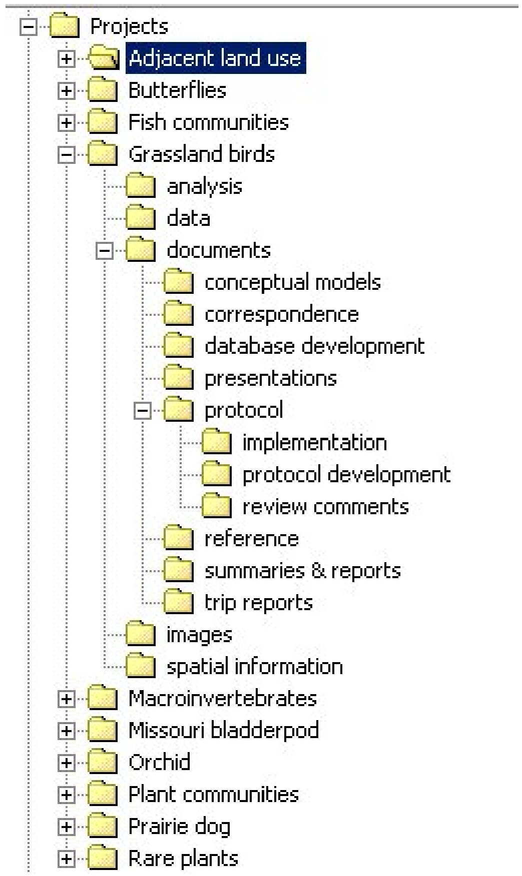
Version Control

Prior to any major changes of a data set a copy is stored with the appropriate version number. This allows for the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Versioning of archived data sets is handled by adding a three digit number to the file name, with the first version being numbered 001. Each additional version is assigned a sequentially higher number. Frequent users of the data are notified of the updates, and provided with a copy of the most recent archived version.

2. Data Organization

The various databases, reports, GIS coverages, etc. used and generated by the monitoring program create a large number of files and folders to manage. Several experiences from the Prairie Cluster reinforce the complicated nature of file management. For example, databases are occasionally stored in two versions of MS Access in order to accommodate data users with different software versions. Further, GIS data are sometimes stored in two projections – one for navigation, the other for use with existing base GIS data. Poor file organization can lead to confusion and data corruption. Figure 8 depicts the file organization structure for the monitoring projects.

Figure 8. File structure for monitoring projects.



E. DATA SECURITY

Project managers maintain the current data files for their project. These data files are stored on the program server. As mentioned in Section I.B.1, the server uses a RAID 5 configuration to ensure data integrity. Access to these files is controlled through the use of password protection. Typically, the Data Manager has universal access to data sets, while access for project managers is limited to their projects. Prior to storage on the server all files are scanned for viruses.

Additional security is provided through tape backups. On a weekly basis the server is backed up using one tape out of a set of 3. One tape holds the current backup, a second holds the previous week's information, and a third tape is stored in the visitor center with information from 2 weeks previous. Personal computers are backed up in a similar manner, but only every two weeks. By using multiple tapes, data from 2-4 weeks past can be restored.

F. DATA SUMMARY, ANALYSIS AND REPORTING

Efficient reporting is important to encourage the use of monitoring data in management decisions. Table 6 outlines the annual reporting schedule for each protocol.

To promote efficient reporting, data management efforts during the summary and analysis phase focus on automation of routine reports. Automation ensures that indices are calculated consistently and removes the potential for error caused by numerous 'cut and paste' operations required in spreadsheets. Furthermore, automated reports provide a consistent format from year to year. Table 7 summarizes our efforts to date towards automating routine reports.

Table 6. Annual schedule for data processing, reporting and archival.

Monitoring Protocol	*Data Collection*	*Data Entry, Verification and Validation*	*Data Analysis/ Reporting*	*Data Archival*
Adjacent Land Use	not applicable			
Grassland Plant Communities	15 Apr – 15 Oct	15 Jan	1 March	1 April
Grassland Birds	1 May – 30 Jun	31 Dec	1 March	15 March
State-listed T&E Plants	1 Apr – 15 Oct	31 Dec		
Missouri Bladderpod (*Lesquerella filiformis*)	1 Sep – 15 June	31 Dec	1 March	15 March
Western Prairie Fringed Orchid (*Platanthera praeclara*)	1 Jul – 31 Jul	31 Dec	1 March	15 March
Black-tailed Prairie Dog (*Cynomis ludovicianus*)	1 May – 15 Aug	31 Dec	1 March	15 March
Local Climate	continuous			
Aquatic Macroinvertebrates	22 May – 19 Sep	1 March	15 April	30 April
Topeka Shiner (*Notropis topeka*)	1 May – 15 Oct	31 Dec	1 March	15 March

Table 7. Long-term schedule for data analysis and reporting.

Monitoring Protocol	*Routine Reporting Interval*	*Tools for analysis*	*Routine Summary Automated?*
Adjacent Land Use	10 years	Imagine	N/A
Grassland Plant Communities	1-2 years	Access, Excel, NCSS	Yes
Grassland Birds	3-5 years	Access	No
State-listed T&E Plants	variable (species-specific)	Arcview, Excel	No
Missouri Bladderpod (*Lesquerella filiformis*)	annual	Access, Arcview, Excel, NCSS	Yes
Western Prairie Fringed Orchid (*Platanthera praeclara*)	Annual	Access, Arcview, Excel, NCSS	No
Black-tailed Prairie Dog (*Cynomis ludovicianus*)	annual	Access, Arcview	No
Local Climate	Annual		N/A
Aquatic Macroinvertebrates	annual	Access, Excel	Yes
Topeka Shiner (*Notropis topeka*)	Annual	Access	No

G. DATA POSTING AND DISSEMINATION

All data collected by the Prairie Cluster is, of course, public property and is subject to requests under the Freedom of Information Act (FOIA). The Channel Island Data Management Protocol describes appropriate procedures to respond to FOIA requests, including the protection of sensitive data such as endangered species locations. In the future, the Prairie Cluster intends to disseminate non-sensitive data through a website. Through the website, those requesting data will be asked to provide information to document by whom and for what purpose the data are being used. By documenting requests, users can be informed when data sets are updated.

H. DATA MANAGEMENT ROLES AND RESPONSIBILITIES

Data acquisition, processing, QA/QC, archiving, and reporting for the various projects are carried out by Prairie Cluster staff members, park natural resource personnel, and by contractors. Each party's responsibility for data management is outlined in Table 8.

Table 8. Responsible parties for various aspects of data management.

Monitoring Protocol	*Data Collection*	*Data Processing*	*QA/QC*	*Analysis/ Reporting*	*Archiving/ Posting*
Adjacent Land Use	Contractor	Contractor	Contractor / Data Mgr.	Data Mgr.	Data Mgr.
Grassland Plant Communities	Project Mgr.	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.
Grassland Birds	Project Mgr.	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.
State-listed T&E Plants	Project Mgr.	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.
Missouri Bladderpod (*Lesquerella filiformis*)	Project Mgr.	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.
Western Prairie Fringed Orchid (*Platanthera praeclara*)	Project Mgr. / Park Staff	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.
Black-tailed Prairie Dog (*Cynomis ludovicianus*)	Project Mgr. / Park Staff	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.
Local Climate	Automated Station	Contractor	Contractor / WIMS	Data Mgr. / Contractor	Contractor / WIMS
Aquatic Macroinvertebrates	Park Staff	Contractor	Contractor / Project Mgr.	Project Mgr.	Data Mgr.
Topeka Shiner (*Notropis topeka*)	Project Mgr.	Project Mgr.	Project Mgr. / Data Mgr.	Project Mgr.	Data Mgr.

III. WORK PLAN

In addition to documenting the data management accomplishments of the Prairie Cluster, the following work plan directs future efforts by defining specific short-term, medium and long-term goals for each of the primary data management objectives. In order for this document to accurately reflect current goals and objectives, the Data Management Plan will be updated every 3-5 years.

Data infrastructure

Short term: **0-1 years**

- Complete database design and data entry forms for all monitoring projects (80% complete).
- Complete transition of legacy data into modern formats (70% complete).
- Finish description of project databases for inclusion in DMP appendices (20% complete).

Medium term: 1-3 years

- Standardize field trip reports.
- Convert all databases to the XP version of MS Access after XP is adopted as the service-wide standard.

Long term: 3-5 years

- Update databases for continued compatibility.

Data integration and exploration

Short term: 0-1 years

- Design prototype graphic user interface, queries and batch commands to integrate data (20% complete).

Medium term: 1-3 years

- Automate routine report generation in MS Access (20% complete).
- Automate import and routine summary of ancillary data sets (e.g. precipitation and stream flow data for macroinvertebrate reporting).
- Integrate all data sets. Promote data integration among I&M Networks in the MWR through the use of standardized core data tables.
- Link spatial data and tabular data using the ArcView to MS Access Link provided by AKSO.

Long term: 3-5 years

- Share data and reports through the I&M web page. Post entire data sets for downloading (with due consideration of security) and distribute reports as PDF files.
- Create a database to track resource management activities and integrate the database with other monitoring data.
- Perform spatial analysis of plant community data.
- Perform spatial analysis of grassland bird data. Incorporate data with regional and national data sets.

Long-term data integrity and security

Short term: 0-1 years

- Update and automate anti-virus software. Implement an off-the-shelf virus software package that will perform automatic scans of all networked computers and download virus updates on a periodic basis (95% complete).
- Update and automate tape backups of both the server and individual machines. Formalize a backup schedule (90% complete).
- Arrange for off-site data storage to guard against a catastrophic event. Schedule regular data shipments to the repository (50% complete).
- Implement data file organization scheme (25% complete).

Medium term: 1-3 years

- Create metadata for spatial data sets and archive data sets on the National Park Service GIS webpage.
- Log data sets into the Dataset Catalog and produce annual updates to NPSpecies.

Long term: 3-5 years

- Update Data Management Plan.

LITERATURE CITED

National Park Service. 1998. Channel Islands National Park, Data Management Protocol. Document available at: http://www.nature.nps.gov/im/units/chis/chisdata/chis.htm.

Tessler, Steven & Joe Gregson. 1997. Draft Data Management Protocol. National Park Service. Document available at: www.nature.nps.gov/im/dmproto/joe40001.htm.

Thomas, Lisa Potter, Michael D. DeBacker, John R. Boetsch, and David G. Peitz. 2001. Conceptual Framework, Monitoring Components and Implementation of a NPS Long-Term Ecological Monitoring Program - Prairie Cluster Prototype Program Status Report. Prairie Cluster Prototype LTEM Program, National Park Service. Document available at: www.nature.nps.gov/im/monitor/prcloverview.pdf.

APPENDIX A. DESCRIPTION OF COMMON DATABASE ELEMENTS

Standardized lookup tables and metadata tables provide consistency and integration among data sets as described in Section II of the Data Management Plan. The design views of shared lookup tables are depicted in Figures A-1 through A-6, and core metadata tables are depicted in Figures A-7 through A-10.

1. SHARED LOOKUP TABLES

Figure A-1. Design of parks lookup table – a comprehensive list of NPS sites.

tbl_Parks_LU : Table

Field Name	Data Type	Description
ParkCode	Text	4-character park code
ParentParkCode	Text	4-character park code pointing to the parent park
ParkName	Text	Park name
Region	Text	NPS region
ParkType	Text	Park type (National Park, National Monument, etc.)

Figure A-2. Design of projects lookup table – information on project supervision, initiation dates, and types of data collected.

tbl_Projects_LU : Table

Field Name	Data Type	Description
ProjectCode	Text	6-character code for the monitoring project
ProjectTitle	Text	Full project title
ProjectManager	Text	Individual in charge of monitoring project implementation
StartDate	Date/Time	When monitoring project was initiated
Comments	Text	Comments on the project
DatabaseName	Text	Name of the data warehouse
Format	Text	Type of database
DatabasePath	Text	Directory path of database
DataTypes	Text	Indicates types of data stored in the database

Figure A-3. Design of species lookup table – a list of species observed and associated taxonomic information.

tbl_Species_LU : Table

Field Name	Data Type	Description
RecordID	AutoNumber	Unique identifier of the record
TSN	Number	ITIS taxonomic serial number
TaxonCode	Text	Locally-unique code for taxa without an assigned TSN
GroupCode	Number	Group of organisms of which this is a member: birds, fish, etc.
ScientificName	Text	Scientific name of the taxon
CommonName	Text	Common or vernacular name for the species
ProvisionalID	Yes/No	Indicates that the species name represents a temporary identification which will be revised once the identity has been better resolved

Figure A-4. Design of staff and cooperators table – a comprehensive list of staff, field workers, and cooperators.

tbl_Staff&Cooperators_LU : Table

Field Name	Data Type	Description
ObserverID	Text	Initials of observer
Name	Text	Full name of observer
Title	Text	Position title of observer
Employer	Text	Employer of observer: default is NPS-PCLTEM
GradeLevel	Text	GS level, or equivalent for non-federal employees

Figure A-5. Design of cover class lookup table – cover classes for estimating percent cover within a defined area.

tbl_CoverClasses_LU : Table

Field Name	Data Type	Description
CoverClass	Text	Cover class for estimating percent cover in a defined area. The program standard is the modified Daubenmire 7-class scale, alternatives are Kelrick's scale described in the original protocol, and North Carolina Vegetation Survey's 10-class scale.
NumericClass	Number	Numeric equivalent of the cover class, used for analysis of log-scaled data
Range	Text	Range of cover values contained by each cover class
MidpointValue	Number	Midpoint of the cover class range
PC-LTEM_CoverClass	Number	Corresponding functional cover classes used to make data compatible with PC-LTEM data collected using the modified Daubenmire 7-class scale
PC-LTEM_Range	Text	Corresponding range of cover values indicated by the functional cover class
PC-LTEM_MidpointValue	Number	Corresponding midpoint of the functional cover class; original class midpoints were used for Kelrick's classes, where the original classes do not line up well with those of PC-LTEM

Figure A-6. Design of habitat attribute lookup table – list of habitat attributes measured for various projects.

tbl_HabitatAttributes_LU : Table

Field Name	Data Type	Description
AttributeID	AutoNumber	Unique identifier for habitat attributes
Attribute	Text	Habitat attribute being measured
AttributeType	Text	Type of attribute being measured: ground cover, species groups, other
Description	Text	Brief description of the habitat attribute

2. CORE METADATA TABLES

Figure A-7. Design of locations metadata table – describes the sites where data collection occurs.

tbl_Locations : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
ParkCode	Text	4-character park code
ProjectCode	Text	6-character code for the monitoring project (Birds, Plant communities, etc.)
Description	Text	Location description
StartUTMX	Number	UTM X (northing) coordinate for the center of the plot or location OR starting point of a line or polygon
StartUTMY	Number	UTM Y (easting) coordinate for the center of the plot or location OR starting point of a line or polygon
StopUTMX	Number	UTM X coordinate (northing) of ending point of line or polygon
StopUTMY	Number	UTM Y coordinate (easting) of ending point of line or polygon
UtmZone	Number	UTM zone
StartLatitude	Number	Latitude in decimal degrees for the center of the plot or location OR starting point of a line or polygon
StartLongitude	Number	Longitude in decimal degrees for the center of the plot or location OR starting point of a line or polygon
StopLatitude	Number	Latitude in decimal degrees for the ending point of a line or polygon
StopLongitude	Number	Longitude in decimal degrees for the ending point of a line or polygon
Datum	Text	Datum of mapping ellipsoid
EstimatedHorizontalError	Number	Estimated horizontal accuracy error--see users guide for complete details and examples
AccuracyComments	Memo	Comments about how positional (horizontal) accuracy was estimated
HabitatType	Text	Habitat type
Elevation	Number	Elevation in meters
Aspect	Number	Slope aspect in degrees (level surfaces have a value of -1)
Slope	Number	Slope angle in degrees
Azimuth	Number	Compass bearing between start and stop coordinates
EstablishmentDate	Date/Time	Date site was established
Discontinued	Date/Time	Date site was discontinued
UnitName	Text	Management unit in which site is located
PLP	Text	Indicates whether location is a point, line, or polygon

Figure A-8. Design of sample periods metadata table – describes the time and duration of each data collection period.

tbl_SamplingPeriods : Table

Field Name	Data Type	Description
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
ParkCode	Text	4-character park code
ProjectCode	Text	6-character code for the monitoring project
StartDate	Date/Time	Date when sampling began
EndDate	Date/Time	Date when sampling ended
TripReport	OLE Object	Trip report in MS Word format, describing protocol implementation details and special circumstances

Figure A-9. Design of observers table – lists the observers for each sampling period.

tbl_Observers : Table

Field Name	Data Type	Description
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
ObserverID	Text	Initials of observer
SamplingHours	Number	Hours spent in the field collecting data, in half-hour increments; excludes travel time
NonsamplingHours	Number	Time spent during the collection period accomplishing project-specific tasks other than field data collection, in half-hour increments; excludes travel time

Figure A-10. Design of sampling events metadata table – optional table describing the time and duration of each data collection event.

tbl_SamplingEvents : Table

Field Name	Data Type	Description
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
EventID	Text	Unique identifier for the sampling event, based on the park code, project code, start date and start time; sampling events are nested within a sampling period
StartTime	Date/Time	Start time
EndTime	Date/Time	End time
StartDate	Date/Time	Date when sampling began
EndDate	Date/Time	Date when sampling ended

APPENDIX B. DESCRIPTION OF THE GRASSLAND PLANT COMMUNITY MONITORING PROTOCOL DATABASE

Title: Grassland Plant Communities	Protocol: Willson, G.D., L.P. Thomas, M.D. DeBacker, W.M. Rizzo and C. Buck. 2001. Plant community monitoring protocol for six prairie parks. Biological Resources Division, U.S. Geological Survey, prepared for Great Plains Prairie Cluster Long-Term Ecological Monitoring Program, Republic, MO.	*Project Manager*: Botanist (Mike DeBacker)
Primary data sets: Frequency and foliar cover of herbaceous and shrub species; number and size of woody species; density of seedlings and saplings *Habitat/Environmental data:* Slope and aspect of sample sites, ground cover of bare soil, exposed rock, leaf litter, etc. *Status:* Implemented at AGFO, EFMO, HOME, PIPE, SCBL, TAPR, WICR *Legacy data sets:*		

Plant Community monitoring has three main components:

- Plant community diversity, composition and structure are monitored through the periodic recording of herbaceous and shrub species occurrences and foliar cover in plots located along permanent transects.
- Density of the woodland overstory is measured using a spherical densiometer, and regeneration of tree species is measured through counts of seedling and saplings
- Habitat variables are important in explaining spatial heterogeneity, and as correlates to trends in the plant community. Slope and aspect are recorded for each plot and the ground cover of bare soil, exposed rock, leaf litter, etc. is estimated.

1. DATABASE STRUCTURE

The plant community monitoring database comprises three primary data tables, two plant attribute tables and core metadata and lookup tables. The occurrence and foliar cover of plant species observed in each plot are recorded in tbl_VegMonData. Canopy cover, measured as the number of points covered by a reflection of canopy vegetation using a spherical densiometer, is recorded in tbl_CanopyCover. Finally, the number of tree seedlings and saplings counted is recorded in tbl_Regeneration. Certain data summaries require species attributes (e.g. native/exotic, shrub/herbaceous plant). Attribute data regarding taxonomy, life history and growth pattern are stored in tbl_PlantAttributes_LU. Further, each species is assigned to a functional guild in tbl_SpeciesWithGuilds for use in summaries (e.g. relative frequency and cover of warm season grasses, spring flowering forbs, etc.). Figure B-1 shows the relationship among tables in the plant community monitoring database. Figures B-2 through B-6 show the fields composing the five protocol specific data tables.

Figure B-1. Plant Community Monitoring Database Structure

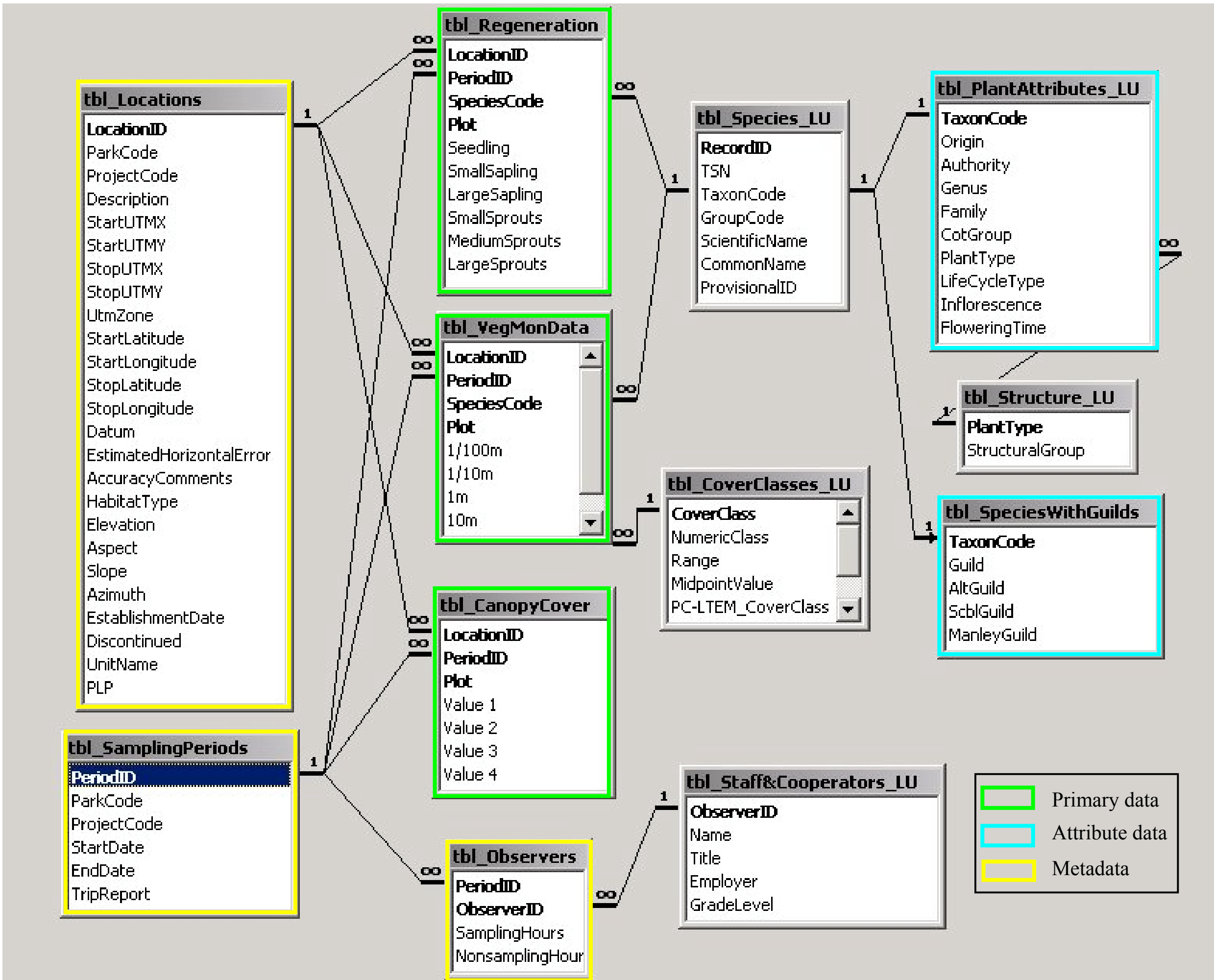


Figure B-2. Design of vegetation monitoring data table – list by site the species observed and their foliar cover.

tbl_VegMonData : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
SpeciesCode	Text	Locally-unique code for taxa without an assigned TSN
Plot	Text	Unique identifier for each plot at that location (usually 10)
1/100m	Number	Presence/absence of species in 1/100m nested subplot. Present = -1
1/10m	Number	Presence/absence of species in 1/10m nested subplot. Present = -1
1m	Number	Presence/absence of species in 1m nested subplot. Present = -1
10m	Number	Presence/absence of species in 10m plot. Present = -1
Cover	Text	Foliar cover class for the species at the 10m scale using Modified Daubenmeier cover classes, values range from 1 to 7

Figure B-3. Design of overstory canopy cover data table – describes the canopy density.

tbl_CanopyCover : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
Plot	Text	Unique identifier for each plot at that location (usually 10)
Value 1	Number	Number of points covered by a reflection of canopy vegetation out of a possible 96 using a sperical densiometer
Value 2	Number	Number of points covered by a reflection of canopy vegetation out of a possible 96 using a sperical densiometer
Value 3	Number	Number of points covered by a reflection of canopy vegetation out of a possible 96 using a sperical densiometer
Value 4	Number	Number of points covered by a reflection of canopy vegetation out of a possible 96 using a sperical densiometer

Figure B-4. Design of the tree regeneration data table – list the number of tree species observed in various size classes.

tbl_Regeneration : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
SpeciesCode	Text	Locally-unique code for taxa without an assigned TSN
Plot	Text	Unique identifier for each plot at that location (usually 10)
Seedling	Number	Tally of seedlings (i.e. < 0.5m tall)
SmallSapling	Number	Tally of small saplings (i.e. > 0.5m tall, < 2.5 cm dbh)
LargeSapling	Number	Tally of large samplings (i.e. > 2.5 cm dbh, < 5.0 cm dbh)
SmallSprouts	Number	Tally of small sprouts (i.e. < 0.5m tall) - sprouts originating from stumps
MediumSprouts	Number	Tally of medium sprouts (i.e. > 0.5m tall, < 2.5 cm dbh) - sprouts originating from stumps
LargeSprouts	Number	Tally of large sprouts (i.e. > 2.5 cm dbh, < 5.0 cm dbh) - sprouts originating from stumps

Figure B-5. Design of plant attributes lookup table – describes the taxonomy, life history and growth patterns of each species.

tbl_PlantAttributes_LU : Table

Field Name	Data Type	Description
TaxonCode	Text	Locally-unique code for taxa without an assigned TSN
Origin	Text	Native/Introduced
Authority	Text	Taxonomic authority
Genus	Text	Taxonomic genus
Family	Text	Taxonomic family
CotGroup	Text	Dicot/Monocot/Gymnosperm/Pteridophyte
PlantType	Text	aquatic/aquatic grass/ aquatic grass-like/fern/grass/grass-like/herb/he
LifeCycleType	Text	Annual/Biennial/AB/AP/BP/Deciduous Perennial/Evergreen Perennial
Inflorescence	Text	Determinate/Indeterminate
FloweringTime	Text	Usual flowering period

Figure B-6. Design of species guilds table – groups all species into functional guilds.

tbl_SpeciesWithGuilds : Table

Field Name	Data Type	
TaxonCode	Text	Locally-unique code for taxa without an assigned TSN
Guild	Text	Primary functional guild
AltGuild	Text	Secondary functional guild
ScblGuild	Text	Guild designations for analysis of Scotts Bluff prairie restoration
ManleyGuild	Text	Guild designations for analysis of Wilson's Creek NB savanna restoration

2. DATA ENTRY

Data entry is initiated through the 'VegDataEntry' form (Figure B-7). For each species in a sample plot, the user selects sample period, location, plot and species name from pick lists. Pick list values are limited to records in the sample periods, locations and species tables. Next, a cover class code is selected from a list of valid numbers, and scale is selected by checking the appropriate box. When entering multiple species from a single plot, the copy button populates the periods, location and plot fields from the previous record. If a species is unknown, the 'species unknown' button opens a form to create a temporary unknown code.

If a species name does not appear in the pick list, either the species is a new occurrence for the park, or an alternative name is used for that species. Synonyms (i.e. two or more different names referring to the same taxon) are common for plant species. The nomenclature update form (Figure B-8), accessed via the 'species synonymy' button, searches for synonyms. Synonymy data are from the USDA PLANTS database. In the form, if the second field contains a common name, the species name is valid and should be added to the species table via the 'enter new species' button (Figure B-9). If the second field begins with an equal sign followed by a different scientific name, then the user should use the new name when entering the species.

Figure B-7. Plant community monitoring data entry form.

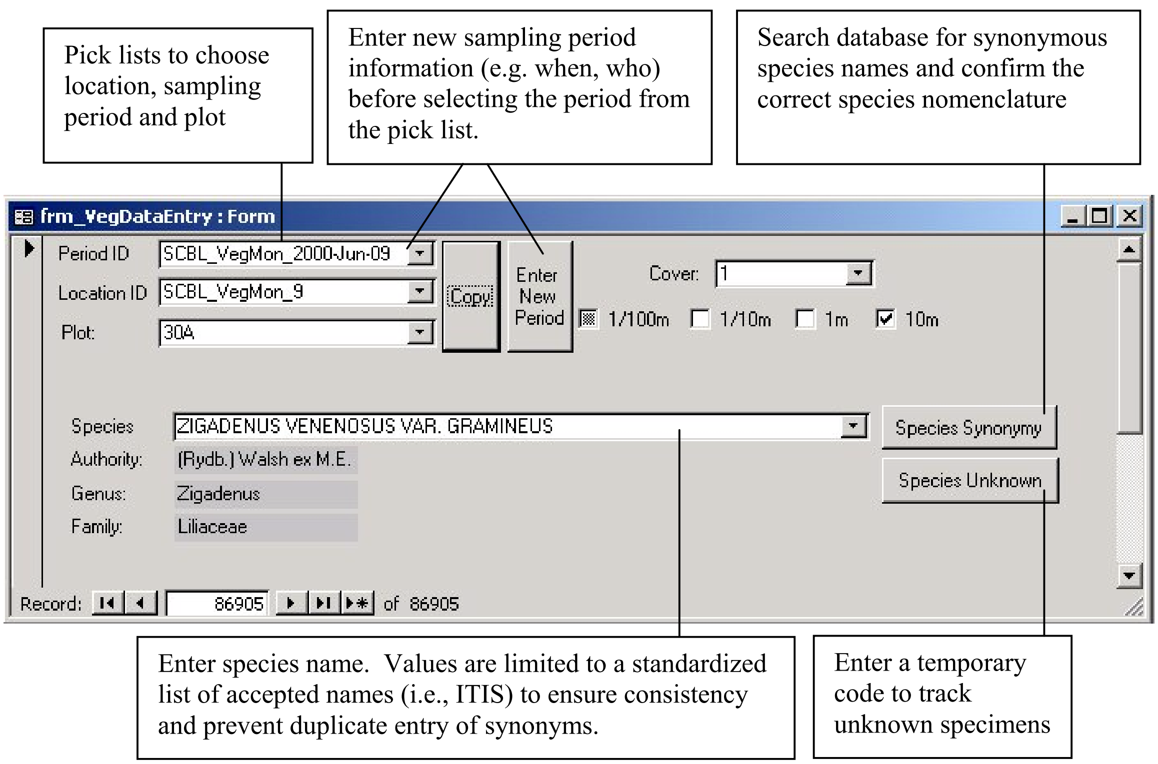


Figure B-8. Species synonymy lookup form.

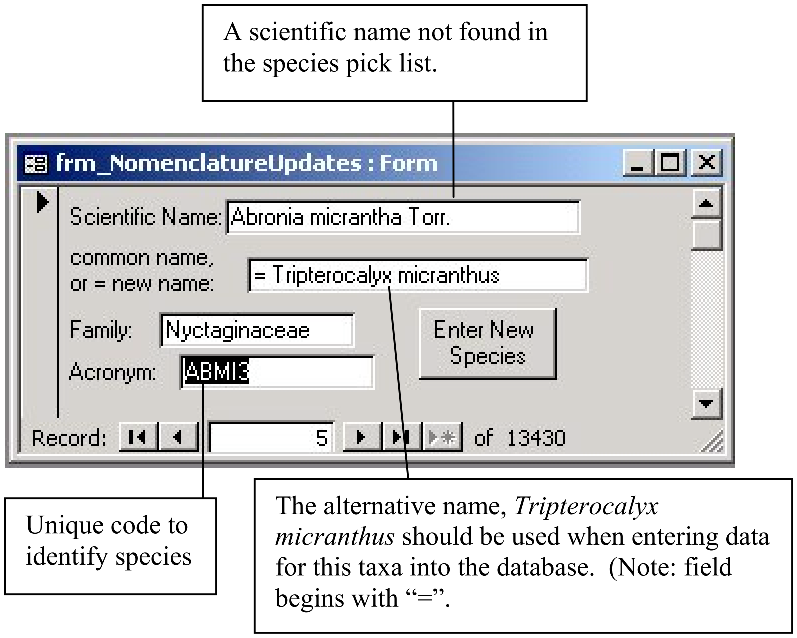
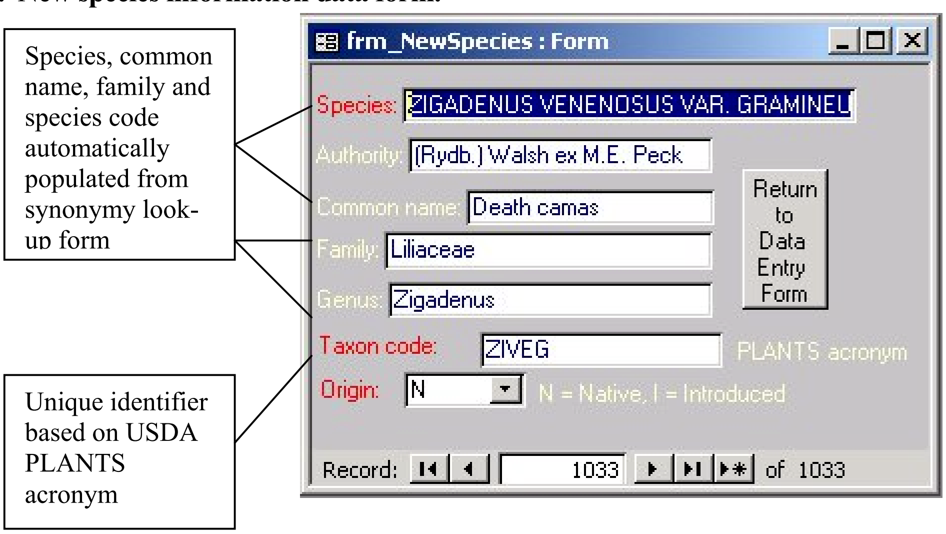


Figure B-9. New species information data form.



3. DATA SUMMARY AND REPORTING

Plant community summary reports are automated in the database, thus eliminating potential error through the many 'cut and paste' operations required when summarizing in a spreadsheet and greatly improving reporting efficiency. Through a series of queries, data from the two sample periods per year are combined to create the following annual reports:

- Plant community composition – Shannon diversity and evenness
- Plant community composition – Relative frequency and cover of plant guilds
- Frequency, mean cover and importance value for all species
- Frequency, mean cover and importance value for exotic species
- Plant community summary – Relative frequency and cover of exotic species
- Plant community structure – Ground cover and vegetation type cover

APPENDIX C. DESCRIPTION OF THE MISSOURI BLADDERPOD MONITORING PROTOCOL DATABASE

<table>
<tr><td>Title:
Missouri Bladderpod
(Lesquerella filiformis)</td><td>Protocol: Kelrick, M.I. 2001. Missouri bladder-pod monitoring protocol for Wilson's Creek National Battlefield. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Missouri Field Station, Columbia, MO. 28 p.</td><td>Project Manager:
Plant Ecologist
(John Boetsch)</td></tr>
<tr><td colspan="3"><u>Primary data sets:</u>
Abundance data used to estimate population size, population demographic data (e.g. individual survivorship and reproductive output).

<u>Habitat/Environmental data:</u>
Slope and aspect, litter and soil depths of demography plots. Estimated cover of associated species, bare soil, exposed rock, leaf litter, etc. See also Local Climate.

<u>Status:</u>
Implemented at WICR

<u>Legacy data sets:</u>
1997-1998 – Michael Kelrick, Truman State University, abundance and demographic data associated with protocol development; 1988-2001 – Prairie Cluster Prototype Program, abundance data collected in a manner different than that specified in the protocol, which was not completed until 2001; 1990-1992 – Lisa Thomas, demographic data collected prior to protocol development</td></tr>
</table>

Missouri bladderpod (*Lesquerella filiformis* Rollins) was listed as Endangered in 1987. Five populations are found at Wilson's Creek National Battlefield. This diminutive winter annual is restricted to limestone glades and rock outcrops in southwestern Missouri and northwestern Arkansas. The habitat structure of the limestone glades has been altered by woody species encroachment, a result of suppression of the periodic wildfires that once maintained their open character. Glade habitat has also been altered and threatened by exotic species establishment.

Monitoring for Missouri bladderpod has three main components:

- Annual abundance estimates are made to determine how population size fluctuates over time. Three different methods for abundance estimation have been implemented for this species: stratified random sampling, adaptive cluster sampling, and comprehensive grid-based censusing. Since 1998, abundance data have been collected within a georeferenced grid established for each site.

- Demographic data are collected periodically from demography plots, which are smaller than the relatively coarse-grained sampling grid cells. Within these plots, individual plants are mapped and tracked through the growing season to determine survivorship and reproductive output. During each return demographic sampling visit, the attributes of surviving individuals (e.g., number of leaves, stems, etc.) are tallied. Seed production is estimated by subsampling mature fruits from surviving individuals. For years with no demographic monitoring, reproductive output is monitored at a more coarse level during abundance sampling by tallying the stems/fruits/etc. in a subsample of plants from each sampled grid cell.

- Habitat data are collected so that patterns of occurrence, survivorship and reproduction can be related to quantified habitat characteristics. The strength and direction of correlation between habitat quality and demographic/abundance patterns is heavily dependent on the scale of observation. For this reason, data are collected at multiple scales so that changes in habitat quality can be monitored at both fine and coarse scales, and cross-scale comparisons can be made.

1. DATABASE STRUCTURE

The database for this monitoring project contains several tables that can be categorized as either metadata tables or primary data tables (Figure C-1). Among the metadata tables for this database are the core metadata tables described in Appendix A (i.e., locations, sampling periods, observers), which contain standardized information about the time and place of data collection. Because demographic data from several sampling periods must be combined in order to summarize survivorship and reproductive output, a project-specific metadata table is available to bind multiple sampling periods to a single sampling season (Figure C-2). Additional project-specific metadata tables describe the configuration of the sampling grids (Figure C-3), and seasonal implementation details for both demographic sampling and abundance estimates (Figures C-4 through C-6). Another metadata table contains file header information needed for automatically building files for importing habitat, abundance and fecundity data to ArcInfo (Figure C-7). Primary data tables contain data that are either related the sampling grid that encompasses each site (Figures C-8 through C-10), or to individual demography plots (Figures C-11 through C-15). Examples of primary data tables include tables for abundance data, habitat data, demographic data, seed production data, and species associate data. An additional primary data table is available for miscellaneous time- and location-specific field observations (Figure C-16). Finally, a single project-specific lookup table is available which contains information about the abundance classes used to estimate the number of individuals in a defined area (Figure C-17).

Figure C-1. Structure of the database for Missouri bladderpod monitoring.

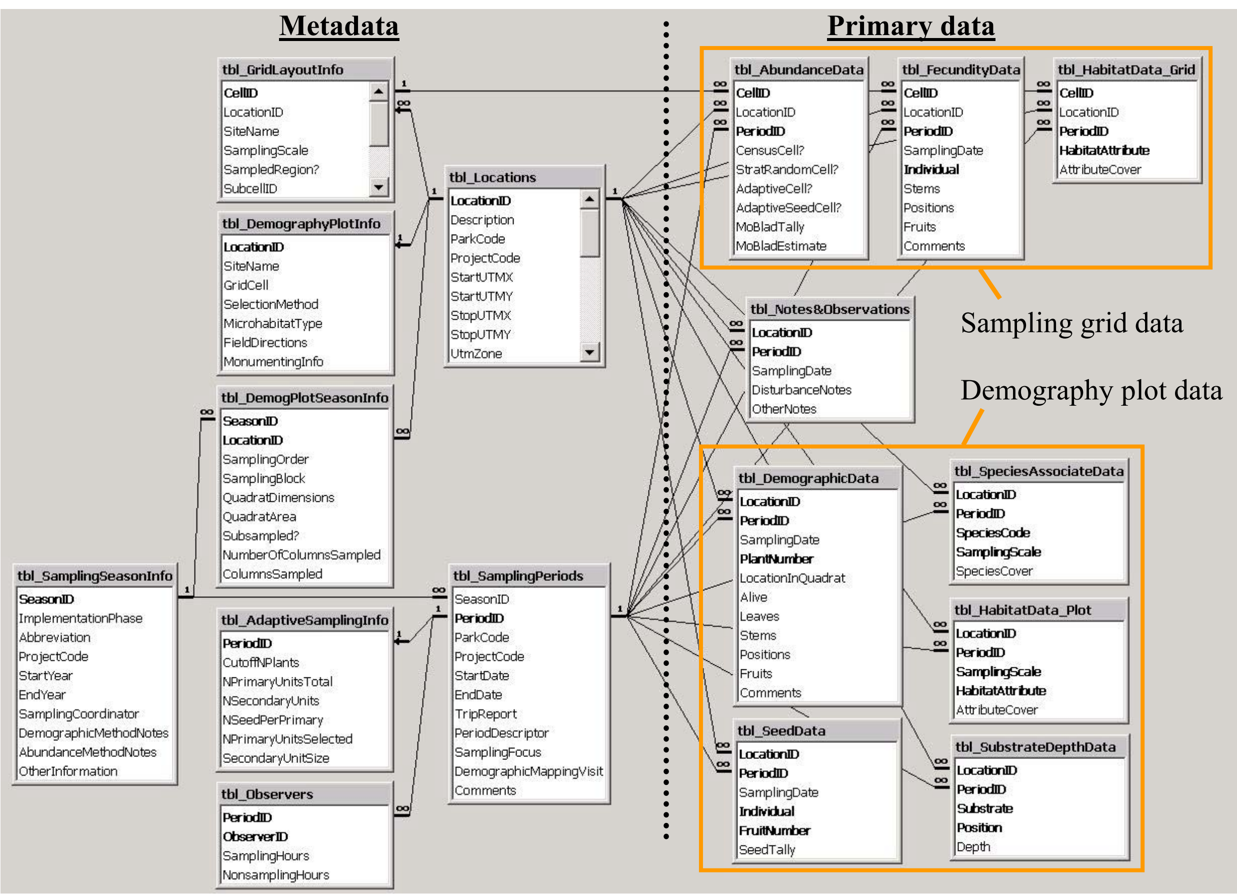


Figure C-2. Design of sampling season metadata table.

tbl_SamplingSeasonInfo : Table

Field Name	Data Type	Description
SeasonID	Text	Unique code identifying the sampling season. Enables grouping of related events to track individuals through a growing season.
ImplementationPhase	Text	Phase of protocol implementation: Legacy Data, Protocol Development, Pilot Study, Full Implementation
Abbreviation	Text	Abbreviated code for this phase of monitoring; this becomes a component of the sampling season ID code
ProjectCode	Text	6-character code for the monitoring project
StartYear	Number	The calendar year in which the sampling season began (typically in autumn)
EndYear	Number	The calendar year in which the sampling season ended (typically in spring)
SamplingCoordinator	Text	Person in charge of implementing the sampling design, data collection, and data analysis
DemographicMethodNotes	Memo	Description of demographic sampling methods, selection criteria, stratification, etc.
AbundanceMethodNotes	Memo	Description of population abundance estimation methods
OtherInformation	Memo	Miscellaneous season-specific information relevant to data interpretation

record-level validation rules:

- the end year of the project is the same or greater than the start year, *or* is null

Figure C-3. Design of grid layout metadata table, containing information about the configuration of the reference grids at each monitoring site.

tbl_GridLayoutInfo : Table

Field Name	Data Type	Description
CellID	Text	Unique grid cell identifier; sorting in ascending order by this field prepares data for GIS export (row by row from left to right)
LocationID	Text	Unique identifier for the location; for BHG, this is the 15-m grid cell; for WnG, this is the 5-m grid cell as the grid at this site was georeferenced at 5 meters
SiteName	Text	Name of the glade site at WICR: BHG = Bloody Hill Glade; NBG = North Bloody Hill Glade; WnG = Walnut Glade; WRG = Wire Road Glade
SamplingScale	Number	Grid cell dimensions: 1-m x 1-m, 3-m x 3-m, or 5-m x 5-m
SampledRegion?	Yes/No	Indicates whether this is in a sampled region of the grid; unsampled regions are primarily filler (NODATA) regions for the GIS and may not be georeferenced
SubcellID	Text	Used to identify the subcell in the field; within the 15-m grid cells at BHG, a-i for 5-m data, and 1-25 for 3-m data
3mID	Text	Used in the field and for data entry to identify the 1-meter grid cell within a given 3-meter subcell (a-i)
5mID	Text	Used in the field and for data entry to identify the 1-meter grid cell within a given 5-meter subcell (1-25)
CoordinateX	Number	Grid coordinates for the NW corner of the cell, in meters distance from the grid origin along the X-axis of the grid
CoordinateY	Number	Grid coordinates for the NW corner of the cell, in meters distance from the grid origin along the Y-axis of the grid

record-level validation rules:

- sampled region is true *and* location ID is not null, *or* sampled region is false

Figure C-4. Design of adaptive sampling metadata table, containing sampling details needed to automate data summarization for two-stage adaptive cluster sampling of abundance.

tbl_AdaptiveSamplingInfo : Table

Field Name	Data Type	Description
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
CutoffNPlants	Number	Minimum number of plants required in a seed cell or sampled member of a cluster to invoke further adaptive sampling
NPrimaryUnitsTotal	Number	Number of possible primary sampling units in the reference frame, within which seed cells are positioned randomly
NSecondaryUnits	Number	Number of secondary sampling units per primary unit (e.g., 25 3-m secondary units within a 15-m x 15-m primary unit)
NSeedPerPrimary	Number	Number of seed cells selected at random within each primary unit (default is 1; database only configured to handle this many at present)
NPrimaryUnitsSelected	Number	The effective sample size, or the number of primary sampling units selected within the reference frame for sampling
SecondaryUnitSize	Number	Dimensions of the secondary sampling unit (default is 3-m x 3-m)

field-level validation rules:
- the cutoff number of plants to invoke adaptive sampling must be greater than 0
- the total number of primary sampling units in the reference frame must be greater than 0
- the number of secondary units in each primary unit must be greater than 0
- the number of seed cells selected in each primary unit must be greater than 0
- the number of primary sampling units selected for sampling must be greater than 0

record-level validation rules:
- the total number of primary sampling units must be greater than the number selected for sampling

Figure C-5. Design of demography plot seasonal metadata table, containing season-specific information about each plot.

tbl_DemogPlotSeasonInfo : Table

Field Name	Data Type	Description
SeasonID	Text	Unique code identifying the sampling season. Enables grouping of related events to track individuals through a growing season.
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
SamplingOrder	Number	Order in which plots were sampled. For BHG, starts from N end of glade and words southward
SamplingBlock	Text	Stratification block. For BHG: 1 = northern third (15-m grid cells 1 - 36); 2 = middle third (cells 37 - 69, 71, 72); 3 = southern third (cells 70, 73 - 107)
QuadratDimensions	Text	Dimensions of demographic quadrats used for this sampling season
QuadratArea	Number	Area of full quadrat for demography plots, in square meters. Typically 50 x 25 cm = 0.125 m2
Subsampled?	Yes/No	Area is subsampled if number of plants in plot is excessive (over 40 plants in a quick count)
NumberOfColumnsSampled	Number	How many of 10 columns were sampled? Each column is 5 cm broad and typically 25 cm long for 50 cm x 25 cm quadrats.
ColumnsSampled	Text	Columns are identified by letter, with column A being the leftmost and column J the rightmost

field-level validation rules:
- the number of columns sampled is greater than 0 and less than or equal to 10, *or* is null

record-level validation rules:
- if the plot was subsampled, the number of columns sampled must be less than 10

Figure C-6. Design of demography plot metadata table.

tbl_DemographyPlotInfo : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
SiteName	Text	Name of the glade site at WICR: BHG = Bloody Hill Glade; NBG = North Bloody Hill Glade; WnG = Walnut Glade; WRG = Wire Road Glade
GridCell	Text	Reference sampling grid cell; at BHG, these cells are 15 x 15 m (the smallest scale at which grid is marked with permanent stakes)
SelectionMethod	Text	How this plot was initially selected: stratified random, subjective placement (where plants occur and/or in specified habitat types), or random subset of subjectively placed plots
MicrohabitatType	Text	Putative habitat type, following the 4-category system defined by Thomas (1996: Natural Areas Journal 16:216-226)
FieldDirections	Memo	Azimuth and distance from rebar stakes at grid corners to plot
MonumentingInfo	Text	Miscellaneous information regarding the hardware in the field used to find the plot

Figure C-7. Design of GIS import metadata table, containing data needed for automatically building files for importing habitat, abundance and fecundity data to ArcInfo.

tbl_GISImportHeader : Table

Field Name	Data Type	Description
SiteName	Text	Name of the glade site at WICR: BHG = Bloody Hill Glade; NBG = North Bloody Hill Glade; WnG = Walnut Glade; WRG = Wire Road Glade
SamplingScale	Number	Grid cell dimensions: 1-m x 1-m, 3-m x 3-m, or 5-m x 5-m
RankOrder	Number	Sort order of the GIS import header
GISImportHeader	Text	Header for importing grid coverages into ArcInfo - habitat, abundance and fecundity data

Figure C-8. Design of abundance data table.

tbl_AbundanceData : Table

Field Name	Data Type	Description
CellID	Text	Unique grid cell identifier; sorting in ascending order by this field prepares data for GIS export (row by row from left to right)
LocationID	Text	Unique identifier for the location; for BHG, this is the 15-m grid cell; for WnG, this is the 5-m grid cell
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
CensusCell?	Yes/No	Is the specified grid cell among those selected for the comprehensive census?
StratRandomCell?	Yes/No	Is the specified grid cell among those selected for stratified random sampling?
AdaptiveCell?	Yes/No	Is the specified grid cell among those selected for adaptive cluster sampling?
AdaptiveSeedCell?	Yes/No	Is the specified grid cell one of the seed cells (randomly-drawn secondary sampling units) for adaptive cluster sampling?
MoBladTally	Number	Tally of individuals within the quadrat or grid cell
MoBladEstimate	Number	Abundance estimate of number of individuals within the quadrat or grid cell

record-level validation rules:

a) the tally field is not null *or* the estimate field is not null
b) *and* census cell is not null *or* stratified random cell is not null *or* adaptive cell is not null
c) *and* the grid cell was either an adaptive cell (True), *or* was not an adaptive cell *and* was not a seed cell

Figure C-9. Design of fecundity data table, containing data related to reproductive output (stems, flowers, fruits) and associated with abundance data.

tbl_FecundityData : Table

Field Name	Data Type	Description
CellID	Text	Unique grid cell identifier; sorting in ascending order by this field prepares data for GIS export (row by row from left to right)
LocationID	Text	Unique identifier for the location; for BHG, this is the 15-m grid cell; for WnG, this is the 5-m grid cell
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
SamplingDate	Date/Time	Date for fecundity sample
Individual	Number	Sequential number identifying each plant
Stems	Number	Tally of stems per individual. NODATA recorded as -1.
Positions	Number	Tally of flower positions per individual. NODATA recorded as -1.
Fruits	Number	Tally of fruits per individual. NODATA recorded as -1.
Comments	Text	Notes regarding plant health and condition

field-level validation rules:
- the number of stems is non-negative, *or* equals –1 (NODATA), *or* is null
- the number of flower positions is non-negative, *or* equals –1 (NODATA), *or* is null
- the number of fruits is non-negative, *or* equals –1 (NODATA), *or* is null

record-level validation rules:
- stems is not null, *or* flower positions is not null, *or* fruits is not null

Figure C-10. Design of habitat data table for grid-based data.

tbl_HabitatData_Grid : Table

Field Name	Data Type	Description
CellID	Text	Unique grid cell identifier; sorting in ascending order by this field prepares data for GIS export (row by row from left to right)
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
HabitatAttribute	Number	Habitat attribute being measured
AttributeCover	Text	Cover value for the habitat attribute

Figure C-11. Design of demographic data table.

tbl_DemographicData : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
SamplingDate	Date/Time	Date plot was sampled
PlantNumber	Number	Sequential number identifying each plant
LocationInQuadrat	Text	For LTEM pilot study, quadrat cells are numbered A-J along the long axis of the 25 cm x 50 cm quadrats, and 1-5 along the short axis
Alive	Yes/No	Indicates whether the individual is alive or dead (missing plants are assumed dead)
Leaves	Number	Tally of leaves per individual. NODATA recorded as -1.
Stems	Number	Tally of stems per individual. NODATA recorded as -1.
Positions	Number	Tally of flower positions (including buds, flowers, fruits and pedicels) per individual. NODATA recorded as -1.
Fruits	Number	Tally of fruits per individual. NODATA recorded as -1.
Comments	Text	Notes regarding plant health and condition

field-level validation rules:
- the number of leaves is non-negative, *or* equals –1 (NODATA), *or* is null
- the number of stems is non-negative, *or* equals –1 (NODATA), *or* is null
- the number of flower positions is non-negative, *or* equals –1 (NODATA), *or* is null
- the number of fruits is non-negative, *or* equals –1 (NODATA), *or* is null

record-level validation rules:
- alive is true *and* leaves is not null *or* stems is not null *or* flower positions is not null *or* fruits is not null, *or* alive is false *and* leaves is null *and* stems is null *and* flower positions is null *and* fruits is null

Figure C-12. Design of seed production data table, containing data for plants subsampled from demography plots.

tbl_SeedData : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period
SamplingDate	Date/Time	Date plot was sampled
Individual	Number	Sequential number identifying each plant
FruitNumber	Number	Sequential number identifying each fruit
SeedTally	Number	Tally of seeds per fruit. NODATA recorded as -1.

field-level validation rules:
- number of seeds per pod is non-negative *and* is less than or equal to 4 (plant develops only 4 ovules per ovary), *or* equals –1 (NODATA)

Figure C-13. Design of species associate data table, containing data on presence and cover for each demography plot.

tbl_SpeciesAssociateData : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
SpeciesCode	Text	Unique identifier for the species
SamplingScale	Number	Scale at which the plot was sampled: quadrat, 1 sq. meter circle, 10 sq. meter circle, left half of quadrat, or right half of quadrat
SpeciesCover	Text	Species cover value

field-level validation rules:
- species cover is non-zero (species that are present must have some cover), *or* equals –1 (NODATA)

Figure C-14. Design of habitat data table for demography plots.

tbl_HabitatData_Plot : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date of the sampling period
SamplingScale	Number	Scale at which the plot was sampled for habitat attribute data: quadrat, 1 sq. meter circle, 10 sq. meter circle, left half of quadrat, or right half of quadrat
HabitatAttribute	Number	Habitat attribute being measured
AttributeCover	Text	Cover value for the habitat attribute

Figure C-15. Design of data table for substrate depths associated with demography plots.

tbl_SubstrateDepthData : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period, based on the park code, project code, and start date
Substrate	Text	Substrate measured: soil or leaf litter
Position	Number	Position at which substrate depth was recorded (positions 1 through 8 specified in the implementation plan)
Depth	Number	Substrate depth in centimeters

field-level validation rules:
- measurement positions are numbered 1 through 8
- depth measurements are non-negative, *or* equal –1 (NODATA)

Figure C-16. Design of data table for field notes and observations (disturbance, phenology, etc.).

tbl_Notes&Observations : Table

Field Name	Data Type	Description
LocationID	Text	Unique identifier for the location, based partly on the park code and project code
PeriodID	Text	Unique identifier for the data collection period
SamplingDate	Date/Time	Date plot was visited for disturbance data
DisturbanceNotes	Memo	Comments on disturbance
OtherNotes	Memo	Other comments on phenology, habitat quality, etc.

Figure C-17. Design of lookup table for the abundance classes used for estimating the number of individuals per area.

tbl_AbundanceClasses_LU : Table

	Field Name	Data Type	Description
🔑	AbundanceClass	Number	Abundance class for estimating the number of individuals in a defined area
	ClassRange	Text	Range of abundance values contained by each abundance class
	LowerBound	Number	Lower bound of the abundance class
	UpperBound	Number	Upper bound of the abundance class

record-level validation rules:

- the upper bound of the abundance class is greater than the lower bound, *or* is null

2. USER INTERFACE

Most routine data handling tasks are managed through a user interface that starts automatically when the database is opened. The user interface is intented to simplify and streamline access to the data, and automate the most frequently needed data products (e.g., summary reports, field data sheets, GIS products). Figure C-18 diagrams the functionality of this user interface. Figure C-19 shows the main switchboard through which the user can access the various data handling modules, examples of which are shown in Figures C-20 and C-21.

Figure C-18. Flowchart of data handling modules within the user interface for the Missouri bladderpod database.

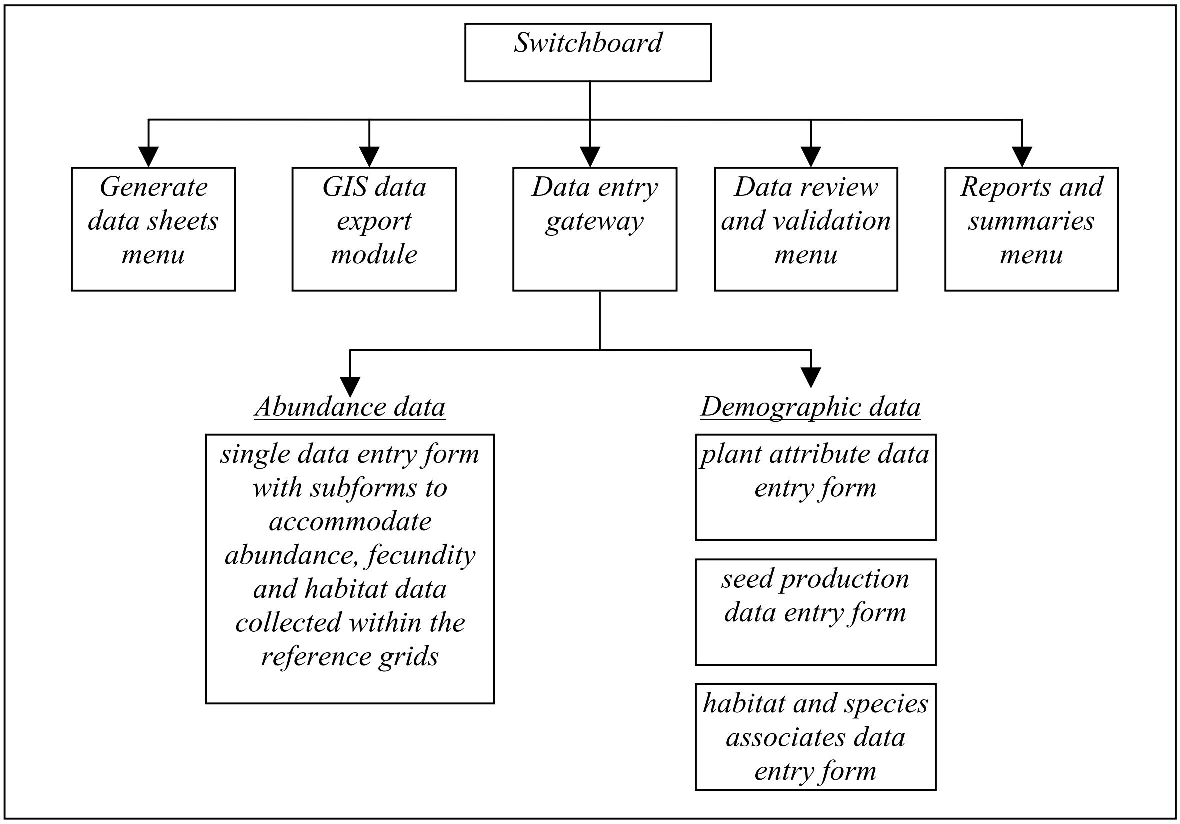


Figure C-19. Main switchboard of the user interface for Missouri bladderpod.

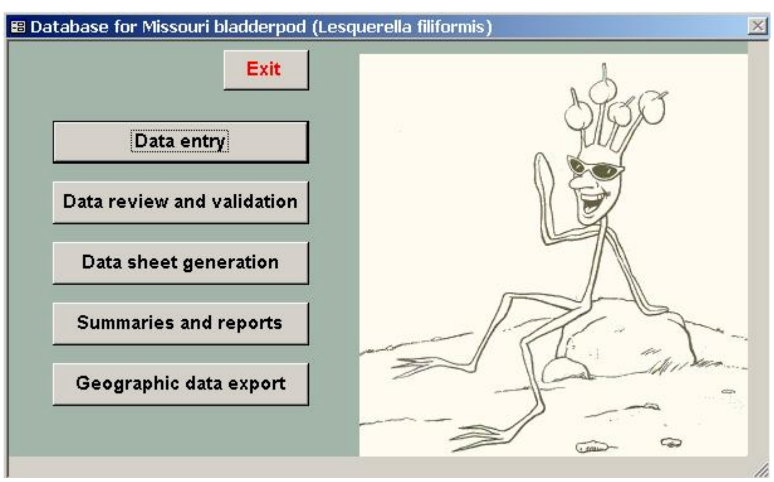
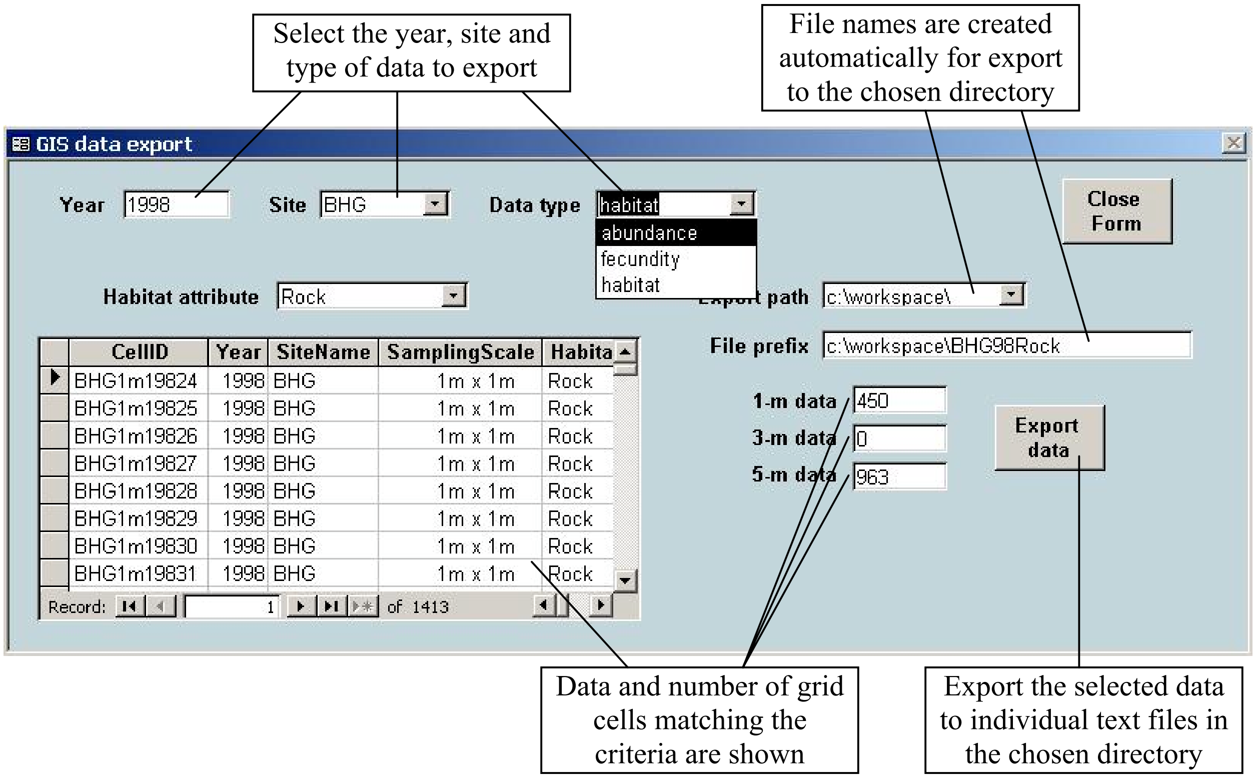


Figure C-20. Interactive module for exporting data to GIS-compatible files.



3. DATA ENTRY

Data entry forms are designed to minimize transcription errors through the use of pick lists, range limits, input masks, and validation rules. Many of these features are built into in the underlying tables, and are documented as part of Figures C-2 through C-17. Referential integrity between related tables is another component which acts to reduce errors by eliminating typographic errors, duplicate records, and orphaned records (i.e., those without supporting information in a related table). Finally, context-sensitive filters help to constrain data based on related information. For example, instead of selecting from the full list of sampling periods, the user first selects the sampling season or year before selecting the appropriate sampling period from a list filtered by the sampling season.

4. DATA SUMMARY AND REPORTING

Summary reports for Missouri bladderpod monitoring are automated through the use of multiple data queries. A user interface has been developed to facilitate report generation (Figure C-21). The user has the option of filtering summary data in an interactive manner by site and/or for a particular year. Automated summaries exist for all aspects of Missouri bladderpod monitoring, and can be categorized as follows:

- Abundance estimate data – comprehensive censusing, adaptive cluster sampling, and stratified random sampling
- Demographic data – survivorship, fecundity (i.e., stems, flowers, fruits), seed production
- Habitat data – species occurrence matrix, species richness, ground covers and species group covers, and substrate depths
- Other – seasonal documentation of protocol implementation, field notes and observations

Figure C-21. Interactive module for generating automated summaries and reports.